

UNITED STATES AIR FORCE  
SUMMER RESEARCH PROGRAM -- 1995  
GRADUATE STUDENT RESEARCH PROGRAM FINAL REPORTS

VOLUME 7A  
ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way  
Culver City, CA 90230-6608

Program Director, RDL  
Gary Moore

Program Manager, AFOSR  
Major David Hart

Program Manager, RDL  
Scott Licoscas

Program Administrator, RDL  
Gwendolyn Smith

Program Administrator, RDL  
Johnetta Thompson

Reproduced From  
Best Available Copy

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH  
Bolling Air Force Base  
Washington, D.C.  
December 1995

19981218 056

## REPORT DOCUMENTATION PAGE

Form Approved

AFRL-SR-BL-TR-98

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing existing information, gathering material and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-018).

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE December, 1995	3. REPORT TYPE Final	0827
4. TITLE AND SUBTITLE USAF Summer Research Program - 1995 Graduate Student Research Program Final Reports, Volume 7A, Armstrong Laboratory			5. FUNDING NUMBERS
6. AUTHORS Gary Moore			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research and Development Labs, Culver City, CA			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NI 4040 Fairfax Dr, Suite 500 Arlington, VA 22203-1613			10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES Contract Number: F49620-93-C-0063			
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release			12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 words) The United States Air Force Graduate Student Research Program (USAF- GSRP) is designed to introduce university, college, and technical institute graduate students to Air Force research. This is accomplished by the graduate students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. Each participant provided a report of their research, and these reports are consolidated into this annual report.			
14. SUBJECT TERMS AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

## **PREFACE**

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 7 is bound in two parts, 7A and 7B. Volume 7A contains #1-18, and Volume 7B contains reports #19-34. The Table of Contents for Volume 6 is included in both parts.

This document is one of a set of 16 volumes describing the 1995 AFOSR Summer Research Program. The following volumes comprise the set:

<b><u>VOLUME</u></b>	<b><u>TITLE</u></b>
1.	Program Management Report
	<i>Summer Faculty Research Program (SFRP) Reports</i>
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A, 5B, & 5C	Wright Laboratory
6A & 6B	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	<i>Graduate Student Research Program (GSRP) Reports</i>
7A & 7B	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10A & 10B	Wright Laboratory
11	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	<i>High School Apprenticeship Program (HSAP) Reports</i>
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A&15B	Wright Laboratory
16	Arnold Engineering Development Center

## **GSRP FINAL REPORT TABLE OF CONTENTS**

**i-xiv**

<b>1. INTRODUCTION</b>	<b>1</b>
<b>2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM</b>	<b>2</b>
<b>3. RECRUITING AND SELECTION</b>	<b>3</b>
<b>4. SITE VISITS</b>	<b>4</b>
<b>5. HBCU/MI PARTICIPATION</b>	<b>4</b>
<b>6. SRP FUNDING SOURCES</b>	<b>5</b>
<b>7. COMPENSATION FOR PARTICIPATIONS</b>	<b>5</b>
<b>8. CONTENTS OF THE 1995 REPORT</b>	<b>6</b>

### **APPENDICIES:**

<b>A. PROGRAM STATISTICAL SUMMARY</b>	<b>A-1</b>
<b>B. SRP EVALUATION RESPONSES</b>	<b>B-1</b>

### **GSRP FINAL REPORTS**



# A STUDY OF ENHANCEMENTS IN CRYOGENIC WASHOUT PROCESS

Leo Bresler  
Graduate Student  
Department of Chemical Engineering

Northwestern University  
2145 Sheridan Road  
Evanston, IL 60208

Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995

# A STUDY OF ENHANCEMENTS IN CRYOGENIC WASHOUT PROCESS

Leo Bresler  
Graduate Student  
Department of Chemical Engineering  
Northwestern University

## **Abstract**

Large rocket motor cryowashout process was studied. Although the process is similar to waterjet applications, the direct use of available models is not possible due to the nature of solid rocket propellant and large temperature difference. A number of enhancements can be implemented within the current installation design to improve the material removal rate and safety of the process. A series of tests should provide additional information needed for further optimization. Possible major design changes needed for overall optimization are also discussed.

# A STUDY OF ENHANCEMENTS IN CRYOGENIC WASHOUT PROCESS

Leo Bresler

## Introduction

US Armed Forces experience an increasing need for environmentally acceptable methods of destruction or conversion of solid rocket propellants. Large quantities of bulk propellants requiring disposal result from propellant and rocket motor manufacturing, management of normal life-cycle support programs, and arms control treaties. The currently used methods of energetic materials disposal, such as open burning, open detonation, and static firing, will have to be replaced due to the recent changes in environmental regulations.

High pressure waterjets have been successfully used to wash out the propellant from 2700 rocket motors. The propellants and their components removed with this process can be reclaimed and reused. The disadvantage of this method lies in generating a secondary waste stream. It consists of water contaminated with soluble components of the propellants such as ammonium perchlorate. To prevent the creation of secondary waste streams in washout a cryogenic fluid (liquid nitrogen) can be used instead of water. Liquid nitrogen is inert, relatively inexpensive, and does not alter chemical and physical properties of the propellant. The prototype cryogenic washout system has been designed and built by General Atomics, Inc.

To advance this process to full scale operations it should be optimized in all possible aspects. The following report will demonstrate means of understanding and optimizing the washout process, provide recommendations for needed changes and further testing required to find optimal operating parameters.

## Modeling of fluid jet cutting

The easiest way to predict the behavior of the washout process is to model it mathematically and to conduct necessary experiments numerically. There have been a number of models developed for water jet cutting processes. The first group of models is based on the assumption of the granular structure of the material being cut. Crow *et. al* (1974) introduced a model based on a following mechanism of material removal; the grains of material experience cavitation drag. The cavitation bubbles are formed due to the pressure difference between the upstream and downstream faces of the grain and the cavitation jet exerts Coulomb friction upon the surface of material. The material is removed due to a combination of friction and pore pressure resulting from fluid permeation into the material. An alternative model, also based on the granular structure assumption, has been proposed by Rehbinder (Rehbinder, 1977 and 1980). According to this model the layers of material are removed due to the pressure of fluid permeated between the grains.

The more universal models are valid not just for granular materials. The Elastic-Plastic response model (Hashish *et. al*, 1978, Hashish *et. al*, 1980) uses Bingham model to describe the time dependent stress-strain relationship of the solid material. It allows linkage of all main parameters of a jet cutting system - standoff distance, jet diameter and pressure, feed rate and material properties with the depth and width of the cut. The latest model has been presented by Hood *et. al* (1990). It is based on modifications of models by Crow and Rehbinder. According to this model the material is removed by ledge breakaway; it assumes that the main force of the jet is exerted on the ledge of material within the kerf.

None of the described models can be directly applied to propellant cryowashout for two main reasons. The first one lies in the resin-like structure of propellants being removed. That means that sufficient experimental data has to be collected to fit the model parameters, or the structure of the model has to be changed to reflect the physical nature of the material. The second reason arises because of substantial temperature difference between the jet and the surface of material. It has been noted (Kiohashi *et. al*, 1978) that this temperature difference significantly changes the parameters of the cutting process. In addition, (as recently found by Whinnery *et. al*, 1995, and Griffiths *et. al*, 1995) the propellants undergo transition to glassy state when cooled to cryogenic temperatures. During this transition the physical characteristics of the material change to a large extent. That results in even larger thermal stresses developing near the jet impact region. The modeling of even static cooling processes requires vast analytical and computational efforts (Whinnery *et. al*, 1995, Griffiths *et. al*, 1995), as well as experimental verification.

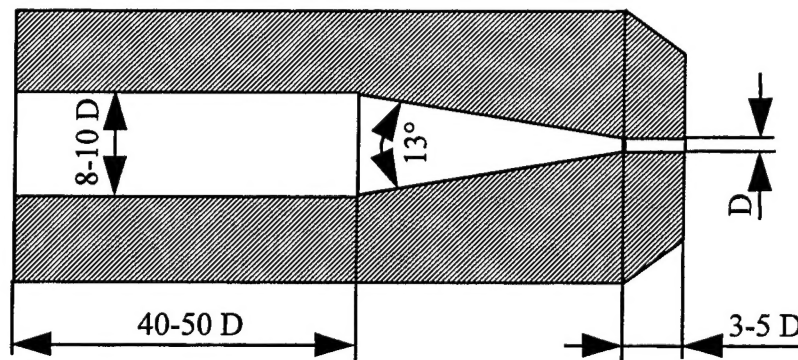
In the following sections we will attempt to show that the enhancements based on simple principles and limited number of experiments may be a cost effective alternative to complete modeling of the problem.

### **Possible enhancements of existing installation**

Fluid jet technology is relatively new. Research in this area has started only about 25 years ago. Although the theory of the processes involved is often complex and not completely understood, the practitioners of fluid cutting have developed a set of process recommendations which can benefit the efficiency of the cryowashout process. The main criterion of the efficiency

is volume of material removed per unit of time. To increase it several improvements can be made.

The first improvement is increasing the depth of cut. To achieve this with minimum modifications, the jet has to operate with minimal losses and close to optimal parameters. To reduce losses some changes can be made in the nozzle design. Studies have shown (Summers, 1995a and references thereafter) that at moderate jet pressures a relatively simple nozzle shape can significantly reduce friction and increase the effective jet length. This “Leach and Walker” design (named after the authors of the 1965 study) (see Fig. 1.) consists of an exit section 3-5 jet diameters long, a linearly expanding section with  $13^\circ$  conical angle, followed by a straight section with diameter 8-10 times that of the jet and 40-50 jet diameters in length. This is close to the current design of the nozzles, so no major changes are required.

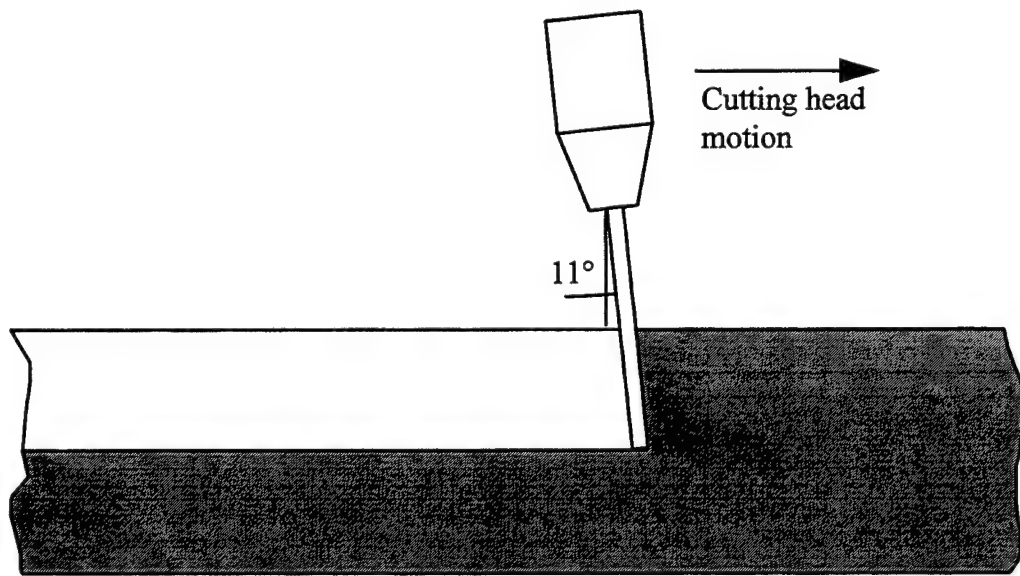


**Figure 1. “Leach and Walker” Nozzle.**

The standoff distance (see Fig. 3) also directly influences the depth of cut. The optimal standoff distance lies within the interval of 30 to 100 jet diameters. In some cases, however, this interval is extended to 120 jet diameters (Summers, 1995a). Due to safety considerations, the

nozzle has to be located as far as possible from the surface of the propellant, consequently the standoff distance should be selected close to the top of the optimal interval.

It has been also noted that the moving jet is being dragged by the friction of material, which reduces the depth of cut. To reduce the drag and increase the depth of cut the jet is inclined in the direction of its motion. The optimal angle is  $11^\circ$  (Summers, 95b), as shown on Figure 2. which shows a view in the traverse plane.

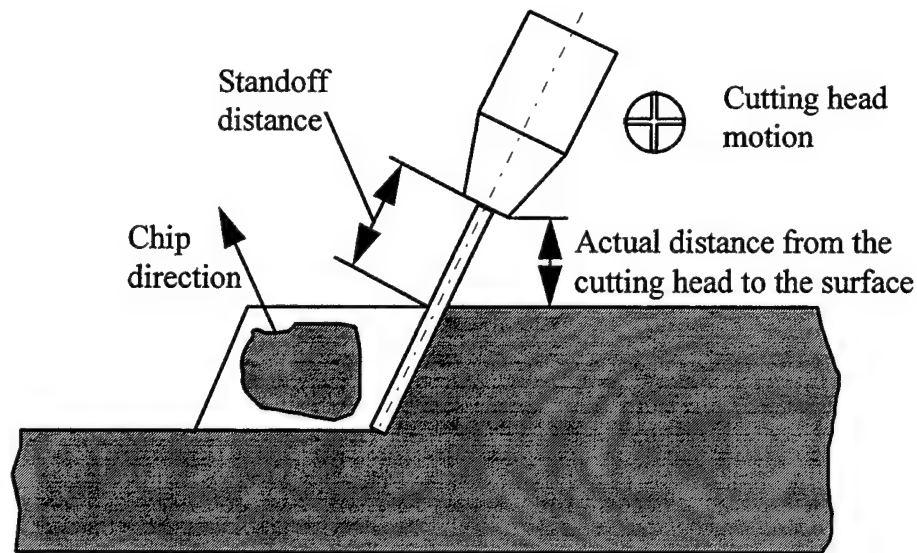


**Figure 2. Optimal jet positioning in the traverse plane.**

Another way to increase the speed of material removal is to use the optimal motion pattern. The cutting head can move along the centerline of the rotating motor. It may also rotate about the lance axis. The combination of these motions can be devised to provide the fastest area coverage and consequently to reduce the time required for the removal of the layer of material.

While increasing the efficiency of the installation, proper attention should be given to safety considerations. Contact between the parts of the cutting system and the energetic material

should be avoided to prevent detonation. The contact can occur between the cutting head and the bulk as well as cut off chips of the propellant. To reduce the possibility of contact several measures can be taken:



**Figure 3. Safe jet positioning.**

To avoid chip reflections in the direction of cutting head, the jet should be inclined in the plane perpendicular to the traverse plane (Fig. 3). In this manner reflected chips will be moving away from the equipment. Although the geometry of the cut will be changed, simple geometric calculation shows that the overall speed of volume removal will not decrease, given that the standoff distance can be maintained. To keep the standoff distance within optimal range, the actual distance between the cutting head and the surface of the bulk material will be reduced, which may increase the possibility of collision between the head and the bulk propellant. Clearly, the optimal angle has to be determined experimentally. In addition, the motion protocol has to be configured to prevent further reduction of the actual distance between the cutting head and the bulk of the propellant.



Many additional enhancements can be made in the design of the installation once the appropriate testing is conducted.

### **Future developments**

To assure optimal performance and safety of the operations a range of tests should be conducted. Most of the tests can be conducted on an inert imitation of the propellant. The following tests will result in determination of parameters for maximum volume removal rate:

- Maximum attainable width of removed layer. A series of cuts needs to be conducted to determine how far from the edge a cut can be made so the chips will still separate from the bulk. This will allow reduction of the distance traveled by the cutting head needed to remove a layer of the material.
- Dependence of the depth of cut on traverse speed. It has been noticed that at relatively high traverse speeds the depth of cut does not significantly reduce with the increase in speed. A series of tests is required to determine a range of traverse speeds optimal with respect to the rate of volume removal.
- Dependence of the depth of cut and maximum width of the chip with the angle of impingement. This will allow determination of an optimal angle of incline in the plane perpendicular to the traverse plane. This optimum may be taken in consideration after the safety conditions are satisfied (see previous section).

The overall changes in installation may allow further increases in the system performance. This increase should be weighted against cost of major changes in design and modifications of the equipment. Possible design changes are:

- Addition of solid abrasive particles ( $\text{CO}_2$ ) into the liquid nitrogen stream. The performance of abrasive fluid jets (depth and width of cut) is considerably higher than that of pure fluid jets. Particles of  $\text{CO}_2$  will not contaminate the removed material or create a waste stream since as liquid nitrogen they will evaporate. This would require major changes in lance and cutting head design as well as in need for  $\text{CO}_2$  source.
- Larger nozzle diameter. This will result in an increase in the depth and width of cut. A higher productivity pump is needed to implement this.
- Cutting head with changing impingement angle. This will allow better material removal near the shell walls with minimal possibility of damage to the casing.
- Additional degrees of freedom for the lance. This will allow better optimization of the removal protocol and increase the volume removal rate.

## **Conclusion**

The present study shows that the cryowashout process can be sufficiently improved with minimal expenses. Testing of the system is required for additional enhancements within the current design. If the changes in equipment become an option, further significant improvements can be made.

## References

- Crow, S.C., Lade, P.V. and Hurlburt, G.H. The Mechanics of Hydraulic Rock Cutting. *Proceedings 2nd Int. Symp. on Jet Cutting Technology*, BHRA, Cranfield, England, paper B1, 1974.
- Griffiths, S., Handrock, J., Kasberg, D., Lipkin, J., Nilson, R., Revelli, V., Weingarten, L., Whinnery, L. Cryocycling of Energetic Materials: Status Report for FY94. *Sandia Report*. SAND95-8235 - UC - 721. Sandia National Laboratories, Livermore, CA, 1995.
- Hashish, M. and duPlessis, M.P. The Applications of a Generalized Jet Cutting Equation. *Proceedings 4th Int. Symp. on Jet Cutting Technology*, BHRA, Cranfield, England, paper F1, 1978.
- Hashish, M. and Reichman, J.M. Analysis of Waterjet Cutting at High Traverse Rates. *Proceedings 5th Int. Symp. on Jet Cutting Technology*, BHRA, Cranfield, England, paper B1, 1980.
- Hood, M., Nordlund, R. and Thimons E. A Study of Rock Erosion using High-Pressure Water Jets. *Int. J. Rock Mech. Min. & Geomech. Abstr.*, **27**, 77-86, Pergamon Press, 1990.
- Kiyohashi, H., Kyo, M., Ishihama, W. Water Jet Breaking of Imitation Hot Dry Rock. *Proceedings 4th Int. Symp. on Jet Cutting Technology*, BHRA, Cranfield, England, paper C2, 1978.
- Rehbinder, G. Slot Cutting in Rock with a High Speed Water Jet. *Int. J. Rock Mech. Min. & Geomech. Abstr.*, **14**, 229-234, Pergamon Press, 1977.
- Rehbinder, G. A Theory About Cutting Rock With a Water Jet. *Rock Mechanics*, **12**, 247-257, Springer-Verlag, 1980.
- Summers, D.A. *Waterjetting Technology*. E&FN Spon, London, England, **1995a**.
- Summers, D.A. *Private Communication*, **1995b**.
- Whinnery, L., Griffiths, S., Lipkin, J., Nilson, R., Goods, S., Dawson, D., Radloff, H., Larson, R., Long, B., Swearingen, J., Hraby, J. Particle Size Reduction of Propellants by Cryocycling. *Sandia Report*. SAND95-8227 - UC - 721. Sandia National Laboratories, Livermore, CA, 1995.

# **A STUDY OF THE MONTE CARLO METHOD**

**Jamylle L. Carter  
Graduate Student  
Department of Mathematics**

**University of California, Los Angeles  
405 Hilgard Avenue  
Los Angeles, California 90025**

**Final Report for  
Graduate Student Research Program  
Armstrong Laboratory**

**Sponsored by  
Air Force Office of Scientific Research  
Bolling Air Force Base  
Washington, DC**

**September 1995**

## A STUDY OF THE MONTE CARLO METHOD

Jamylle L. Carter  
Graduate Student  
Department of Mathematics  
University of California, Los Angeles

### Abstract

The Monte Carlo method was studied and applied to the integral and the heat conduction equation. For the integral, the error of the Monte Carlo method was estimated and a precise mathematical explanation given for its use. Results were satisfactory for the integral yet intriguing for the heat equation. The Monte Carlo heat equation solution, derived from a finite-difference equation, suggests either the presence of another mechanism in the physical experiment, i.e., that the original analytical equation may contain error, or the need for greater discretization in the Monte Carlo code.

# A STUDY OF THE MONTE CARLO METHOD

Jamylle L. Carter

## Introduction.

The Mathematical Products Division of the Armstrong Laboratory at Brooks Air Force Base, San Antonio, Texas, has been investigating the use of stochastic or random processes to develop an equation solver for various forms of the wave equation as derived from Maxwell's equations. Investigators in this division are studying this approach in an attempt to develop a robust numerical tool for computing the electromagnetic fields that propagate into the human body from various external sources. This effort has led to considerations of the Feynman integral method and Monte Carlo methods. This report begins an evaluation of Monte Carlo methods applied to partial differential equations to include, ultimately, the wave equation. Some basic principles of Monte Carlo are reviewed for the prototypical problem of solving integrals, and a study of a heat diffusion problem over a finite line is described.

## Discussion of Problem.

What exactly is the Monte Carlo method? Simply put, it is a method for approximating the solution of a system of equations. Difficult problems often necessitate a numerical approximation rather than an analytical solution. To get a numerical approximation using the Monte Carlo method, we must first make a model of the system (Rubinstein, page 11). The Monte Carlo method involves running several experiments on that model and then analyzing the results. The results of the analysis are approximately equal to the solution of the system. In most cases these results estimate the expectation of the random variable associated with the deterministic problem. In other words, with the Monte Carlo method we try to construct an experiment which accurately models the system, a random variable whose expectation is the solution of the system, and a suitable approximation of that expected value.

The Monte Carlo method is best suited for problems which can be easily simulated by a random process. Two such problems are the definite integral and the heat equation. The definite integral was particularly studied because the Mathematical Products Division is interested in solving Maxwell's equations, which can be represented in integral form. The heat conduction equation was studied to gain insight into the Monte Carlo method by a more traditional approach.

A statistical analysis yields the error of the Monte Carlo method. The goal is to construct both an accurate experiment and a random variable whose expected value is our unknown solution: the Monte Carlo method will then approximate this expected value. Suppose we have an experiment such that the probability of some event  $B$  equals the solution of the system: call this probability  $p$ . In practice  $p$  would be unknown before the experiment. We need to construct a random variable with expectation  $p$ , also. Consider the Bernoulli random variable  $\psi_i$  which takes the value 1 if the event  $B$  occurs in the  $i$ th experiment and zero otherwise. The expected value of  $\psi_i$  is  $p$  since  $\psi_i$  relates to the event  $B$ . How can we approximate  $p$ , the solution of the system? We can approximate  $p$  by the frequency of the occurrence of the event  $B$  (Papoulis, page 8). If we run  $N$  experiments, the sum over all  $i$  of the  $\psi_i$ 's gives the number of times that  $B$  occurred in the  $N$  experiments. If we call this sum the random variable  $M$  with observed value  $M$ , then the frequency of the event  $B$  can be written as  $M/N$ , where  $M/N$  approximates the true probability  $p$  of  $B$ . But  $M/N$  is the observed value of the random variable  $M/N$ , whose expectation is  $p$  by construction. So  $M/N$  approximates  $p$ , the expected value of the random variable  $M/N$  which we symbolize as  $p$ . Since  $p$  is the solution we are seeking and  $M/N$  its estimate, then the error of the Monte Carlo method would be the absolute value of  $M/N - p$ , written  $\text{abs}(M/N - p)$ .

But the error  $\text{abs}(M/N - p)$  appears in a famous inequality, the Chebyshev inequality.  $M/N$  has mean  $p$  and variance  $p(1 - p)/N$  since this ratio is the result of Bernoulli trials. Chebyshev's Inequality gives an upper bound on the probability that a random variable takes its values away from the mean:

$$\text{Probability}(|\frac{M}{N} - p| \geq k\sigma) \leq \frac{1}{k^2}$$

In this equation  $\sigma$  is the standard deviation (the square root of the variance) of the random variable in question. Since in our case

$$\sigma = \sqrt{p(1 - p) / N}$$

it is seen that the error of the Monte Carlo method is of the order  $1/\sqrt{N}$ . Thus we can reduce the error in the Monte Carlo method by increasing the number of experiments.

To approximate a one-dimensional integral by the Monte Carlo method, we first must construct the right experiment. One of the simplest models is given by the "Hit or Miss" Technique. This technique is based on interpreting the integral as the area under the curve. Suppose that we are given a non-negative bounded function  $f(x)$  over an interval  $[a, b]$ ; say  $f(x)$  is bounded above by  $c$ . The integral of  $f(x)$  over  $[a, b]$  can be thought of as the area between the  $x$ -axis and the curve  $f(x)$ . Certainly this area is less than or equal to the area of the bounding rectangle  $c(b-a)$ , so their ratio is less than or equal to 1. The random selection of points in the rectangle will approximate this area ratio. If the point falls below the curve, it is labeled a hit; otherwise it is called a miss. The ratio of the number of hits (the number of points beneath the curve) to the total number of points is almost equal to the area ratio. If we multiply this *hit/total* ratio by the known rectangle area, then we get an estimate of the integral.

We can describe this experiment in more precise statistical terms. The solution of the system, i.e., the value of the definite integral, is the probability that a random point  $(\psi, \nu)$  uniformly distributed in the rectangle  $[a, b] \times [0, c]$  lies below the curve  $f(x)$ . The selection of this random point is our experiment, and the previously cited event  $B$  occurs when the point falls below  $f(x)$ . Our Bernoulli random variable is  $\chi(\psi, \nu)$  where the function  $\chi(\psi, \nu)$  takes the value 1 if  $\nu < f(\psi)$  and 0 otherwise (Shrieder, pages 92-96).

In the following, for clarification, a random variable  $z$  is written in boldface and its observed value  $z$  in italics. The expected value of  $g(\psi, \nu)$  is determined by the following theorem: If  $g(\mathbf{x}, \mathbf{y})$  is a function of two random variables  $\mathbf{x}$  and  $\mathbf{y}$  with joint density  $h(\mathbf{x}, \mathbf{y})$ , then the expectation of  $g(\mathbf{x}, \mathbf{y})$  is equal to the double integral of the product of  $g(\mathbf{x}, \mathbf{y})$  and  $h(\mathbf{x}, \mathbf{y})$ , taken from  $-\infty$  to  $+\infty$  (Papoulis, page 206). The joint density  $h(\psi, \nu)$  of the random variables  $\psi$  and  $\nu$  is  $1/[c(b-a)]$  if  $(\psi, \nu)$  lies in the rectangle of integration and 0 otherwise. We have

$$\text{Expectation}(\chi(\psi, \nu)) = \int_a^b \int_0^c \frac{1}{c(b-a)} \chi(x, y) dx dy$$



In the above expression

$$\begin{aligned}\chi(x, y) &= 1 \text{ iff } y \leq f(x) \\ \chi(x, y) &= 0 \text{ iff } y > f(x)\end{aligned}$$

Expectation  $\chi(\psi, v)$  is the ratio of the integral area to the rectangular area. This can be seen as follows:

$$\text{Expectation}(\chi(\psi, v)) = \int_a^b \int_0^c \frac{1}{c(b-a)} \chi(x, y) dx dy$$

$$= \frac{1}{c(b-a)} \int_a^b \left[ \int_0^c \chi(x, y) dy \right] dx$$

$$= \frac{1}{c(b-a)} \int_a^b \left[ \int_0^{f(x)} dy \right] dx$$

$$= \frac{1}{c(b-a)} \int_a^b f(x) dx$$

Then if we define the random variable  $S$  as  $S := c(b-a)\chi(\psi, v)$ , the expectation of  $S$  is just the integral itself. We can estimate this expectation by choosing  $N$  random number pairs  $(\psi, v)$  with  $\psi$  drawn from a uniform distribution with density  $1/c$  and  $v$  drawn from a uniform distribution with density  $1/(b-a)$  and counting the number of times that  $\psi \leq f(v)$ . (Shrieder, page 96).

The Monte Carlo method allows us to compute the definite integral of any bounded function over a bounded domain, no matter how complicated the integrand. Using the Chebyshev Inequality we see that the error of the Monte Carlo method is still of the order  $1/\sqrt{N}$ . Consider an example: Suppose that we want to calculate the integral of  $e^x$  over the interval  $[0, 1]$ . (A well-known integral is chosen to illustrate the convergence of the method.) Over  $[0, 1]$   $e^x$  is bounded above by 3, so the area of the bounding rectangle is 3. (The choice of the x-axis as the lower bound of the rectangle will not affect the value of the approximation.) We will approximate integral area/rectangle area by (the number of random points falling below  $e^x$ )/the total number of points. If we multiply the latter ratio by the rectangle area 3, then we get an estimate of the integral. MATLAB code which implements the Hit or Miss Monte Carlo method for the integral is shown in the methodology section.

The heat equation is a tool for studying the Monte Carlo method which provides a good entry into the problem of partial differential equation solvers. Consider applying a constant heat source to one end of the unit bar and insulating the other end; suppose the bar has zero initial temperature. The heat equation which represents this experiment is as follows:

$$\frac{\partial v}{\partial t} = k \frac{\partial^2 v}{\partial x^2}$$

In this equation  $v(x,t)$  = temperature at point  $x$  and time  $t$ ,  $k$  is a positive constant, and  $v(x,0) = 0$  for every  $x$  on the finite line from 0 to  $D$ . We take

$$\frac{\partial v}{\partial x} = 0$$

at  $x = D$  and

$$\frac{\partial v}{\partial x} = -Q$$

at  $x = 0$ .

The conduction of heat is a random diffusive process, so it is perfectly suited for a Monte Carlo approximation. The method we used discretizes space and time. Discretize time by  $t_j = j(\Delta t)$  where  $\Delta t = 1$  second and space by  $s$  subintervals (or bins) of equal length. The discrete function  $n(i, j)$ , which represents the number of "heat quanta" in the  $i$ th subinterval at time  $j$ , will approximate  $v(x_i, t_j)$  for fixed values of  $t$  and  $x$ . Simulate the influx of heat by putting one new heat molecule into the bar each second. In the actual experiment, heat makes the bar's molecules move around quickly. We model this movement by assuming that in each second each particle has the probability  $p$  of moving left and  $p$  of moving right; thus the probability of moving in either direction is  $2p$ , forcing  $p \leq \frac{1}{2}$ . In each second, a random number  $x$  in  $[0,1]$  is assigned to each molecule. If  $x$  is less than or equal to  $2p$ , then the particle moves. (Note that the probability of  $x$  being less than  $2p$  is exactly  $2p$  since  $x$  is uniformly distributed in  $[0,1]$ , so this random number generation actually represents the probability that a particle moves.) Whether it moves left or right depends on the value of another random number  $y$  in  $[0,1]$ :

$y < \frac{1}{2} \Rightarrow$  molecule moves left

$y \geq \frac{1}{2} \Rightarrow$  molecule moves right.

So in the  $j$ th second we evaluate  $n(i, j)$  for all the subintervals of  $[0,1]$ ; thus we get an approximation for the value of  $v(x, t)$ . We do these computations for all values of  $t$ . (Here  $t$  is a scalar value, not a variable.)

How does this method model a finite difference equation?  $\partial v / \partial t$  is approximated by  $\Delta n(i, j) / \Delta t =$  net change in  $n(i, j)$  during time interval  $\Delta t$ . Let  $p =$  probability that a particle in bin  $i$  moves left to bin  $(i-1)$  at any time  $j$  and let  $p =$  probability that a particle in bin  $i$  moves right to bin  $(i+1)$  at any time  $j$ . Then

$$\begin{aligned}
 \partial v / \partial t &\approx \Delta n(i, j) / \Delta t \\
 &= + p * n(i-1, j) && \text{Gain from bin } i-1 \\
 &\quad - p * n(i, j) && \text{Loss to bin } i-1 \\
 &\quad + p * n(i+1, j) && \text{Gain from bin } i+1 \\
 &\quad - p * n(i, j) && \text{Loss to bin } i+1 \\
 &= p * n(i-1, j) - 2p * n(i, j) + p * n(i+1, j) \\
 &= p * [n(i-1, j) - 2 * n(i, j) + n(i+1, j)] \\
 &\approx p * \partial^2 v / \partial x^2
 \end{aligned}$$

So according to the model,  $n(i, j)$  also satisfies the centered finite difference scheme for the second-order spatial derivative.

The MATLAB code which implements the Monte Carlo solution is shown in the methodology section.

### Methodology.

The MATLAB code that solves the integral equation is displayed below:

below = 0;	below = number of points below the curve $\exp(x)$
count = 0;	count = number of points in the rectangle $[0, 1] \times [0, 3]$
for i = 1:1000	For each of the 1000 randomly selected points in the rectangle $[0, 1] \times [0, 3]$
psi = rand(1);	Choose a random number $\psi$ in $[0, 1]$ , the $x$ -value of point.
clear rand(1);	
nu = 3*rand(1);	Choose another random number $\nu$ in $[0, 3]$ , the $y$ -value of the point.
if nu < exp(psi)	If the random point falls below the curve $\exp(x)$ ,
below = below + 1;	increment the number of points below $\exp(x)$ .
end	
count = count + 1;	Increment the number of points in the rectangle.
frac = below/count;	frac = $\frac{\text{number of points falling below } e^x}{\text{total number of points}}$
int(i) = 3*frac	int(i) approximates the value of the integral after the $i$ th point.
error(i) = abs(exp(1) - 1 - int(i));	Error after $i$ th point =   exact solution - approximate solution
end	
for i = 1:1000	
i, int(i), error(i)	Prints the value of the approximated integral and its

end                                      error after the  $i$ th point.

The MATLAB program for the Monte Carlo computation involving the heat conduction equation is given below:

% This program is a Monte Carlo method for solving the one-dimensional heat  
% equation.

% If  $v(x,t)$  = temperature at point  $x$  and time  $t$ ,  
% then the heat equation is given by  $p \cdot d^2(v)/dx^2 = dv/dt$ , where  $p$  is a  
% constant.

% Our initial condition is zero temperature in the bar  $[0,1]$ :

%          $v(x,0) = 0$ .

% Our boundary conditions are flux of 1 in from the left, no flux from  
% the right:  $dv/dx = -1$  at  $x=0$

%          $dv/dx = 0$  at  $x=1$

% We approximate the solution  $v(x,t)$  by

% a discretized form of  $v(x,t)$ . We discretize time into  $t$  seconds--

% the user chooses the value of  $t$ --

% and we discretize the interval  $[0,1]$  by subdividing it into

% subintervals (or bins).

% By inputting 1 new molecule

% each second on the left end of the bar, we determine the flux at  $x=0$ .

% We approximate  $d^2(v)/dx^2$  by a centered finite difference scheme.

% Our discretized function is  $n(i)$  = number of heat particles in the  $i$ th bin

% at some specified time  $t$ .

% We describe the discretized function by the following finite difference  
% model:

% Assume that during each second, each particle has an equal probability  $p$  of  
% moving left or right.

%  $p = \text{prob}(1 \text{ particle in bin } i \text{ moves left to bin } i-1)$

%     $= \text{prob}(1 \text{ particle in bin } i \text{ moves right to bin } i+1)$

% Then during one second, the net change of particles in bin I equals

```
% p*n(i-1)      ( bin i-1 -> bin i)
% - p*n(i)      ( bin i-1 <- bin i)
% - p*n(i)      ( bin i  -> bin i+1)
% + p*n(i+1).   ( bin i  <- bin i+1)
```

% So net change after 1s =  $p \cdot n(i-1) - 2p \cdot n(i) + p \cdot n(i+1)$

% =  $p \cdot [n(i-1) - 2 \cdot n(i) + n(i+1)]$

% ~  $p \cdot d^2(v)/dx^2$

% We solve this difference equation for each second.

% At time t the values of n(i)

% give the distribution of particles in the bar after t seconds.

% We need to define inc(i) = increment of particles in bin I during

% 1s period.

%-----

% Discretize time into t 1-second intervals.

t = input('How long? ')

% Since p = prob moving left

% = prob moving right

% then 2p = prob moving at all

% so  $0 \leq 2p \leq 1$ .

% In other words,  $p \leq 0.5$

p = input('What probability? p must be  $\leq 0.5$  ')

% Testing the code with small number of subintervals.

s = input('Number of subintervals? ')

n = zeros(1,s);

inc = zeros(1,s);

```
for j = 1:t
```

```
    % n(1) -> n(2)
```

```
    for k = 1:n(1)
```

```
        if rand(1) < p
```

```
            inc(1) = inc(1) - 1;
```

```
            inc(2) = inc(2) + 1;
```

```
        end
```

```
        clear rand(1)
```

```
    end
```

```
    % n(1) <- n(2) -> n(3)
```

```
    for k = 1:n(2)
```

```
        if rand(1) < 2*p
```

```
            clear rand(1);
```

```
            y = rand(1);
```

```
            if y < 0.5
```

```
                inc(1) = inc(1) + 1;
```

```
                inc(2) = inc(2) - 1;
```

```
            else
```

```
                inc(3) = inc(3) + 1;
```

```
                inc(2) = inc(2) - 1;
```

```
            end
```

```
        end
```

```
        clear y rand(1)
```

```
    end
```

```
    inc(1) = inc(1) + 1;
```

```
    % n(i-1) <- n(i) -> n(i+1)
```

```
    for i = 3:(s-1)
```

```

for k = 1:n(i)
    if rand(1) < 2*p
        clear rand(1);
        y = rand(1);
        if y < 0.5
            inc(i-1) = inc(i-1) + 1;
            inc(i) = inc(i) - 1;
        else
            inc(i+1) = inc(i+1) + 1;
            inc(i) = inc(i) - 1;
        end
    end
    clear y rand(1)
end
end

```

```

% n(s-1) <- n(s)

```

```

for k = 1:n(s)
    if rand(1) <= p
        inc(s-1) = inc(s-1) + 1;
        inc(s) = inc(s) - 1;
    end
    clear rand(1)
end

```

```

for i = 1:s
    n(i) = n(i) + inc(i);
end

```

```

inc = zeros(1,s);

```

```

end

```



## Results.

The integral of  $e^x$  from 0 to 1.0 is 1.718. The Monte Carlo result we obtained by running the first of the MATLAB programs provided in the methodology section above was 1.677. Since 1000 random draws were used in the Monte Carlo solution, the Monte Carlo result is of the order  $(1/\sqrt{1000})$ .

The Monte Carlo program was run using only five spatial bins. However, the results are in qualitative agreement with an analytical solution that we developed and with an analytical solution available in Carslaw and Jaeger (Carslaw and Jaeger, page 112). Figures 1 through 4 show typical results. The ordinate of these figures displays the number of heat quanta. For example, in the figure labeled time = 10, 10 "particles" or quanta have been released into the discretized line segment. What is seen by reviewing Figures 1 through 4 is the penetration of heat into the line and the slow build up of average temperature (proportional to heat content) since no heat loss is allowed.

Referring again to the heat conduction problem being solved, the analytical solution that was developed is displayed below. Recall that the basic heat conduction equation is:

$$\frac{\partial v}{\partial t} = k \frac{\partial^2 v}{\partial x^2}$$

and that we are solving this equation under the conditions that  $v(x,0) = 0$  for every  $x$ , and

$$\frac{\partial v}{\partial x} = 0$$

at  $x = D$ , and

$$\frac{\partial v}{\partial x} = -Q$$

at  $x = 0$ . The analytical solution we have developed for this problem is:

$$v(x, t) = \frac{Qx^2}{2D} - \frac{QD^2}{3} - Q(x-D) + \frac{Qt}{D} + \frac{QD^2}{\pi n} \sum_n e^{\frac{-4\pi^2 n^2 kt}{D^2}} \left[ \left( \frac{1}{(2\pi n)^2} + \frac{1}{2} - \frac{1}{D} \right) \sin \frac{2\pi nx}{D} + \left( \frac{1}{2\pi n} \right) \cos \frac{2\pi nx}{D} \right]$$

In our research thus far we have not been able to show satisfactory agreement between the analytical solution given above and the Monte Carlo simulation. The analytical solution shows very rapid convergence to steady state behavior and, in the steady state, is a factor of 2.5 from the Monte Carlo estimates at least. Three avenues of further work to resolve this discrepancy are possible. First, disagreement between the Monte Carlo solution and the analytical solution may be due to the extremely coarse spatial grid used in the Monte Carlo solution, and this can be investigated by increasing the spatial grid in the MATLAB program. Second, an in-depth review of our analytical solution might disclose an error in the lengthy derivation. Finally, further consideration needs to be given to whether the analytical solution obtained conforms to the physics of the Monte Carlo study. Specifically, are the analytical boundary conditions correct?

### Conclusion.

We found it easy to write Monte Carlo programs that appealed to our physical intuition. Both programs provided results which similarly agreed with our physical intuition. An analytical solution we developed to correspond to the Monte Carlo result for the heat equation provided results qualitatively similar to the Monte Carlo solution, but significant quantitative differences were noted. Resolution of this discrepancy awaits further work.

The Monte Carlo method appears to be a meaningful adjunctive method in the solution of partial differential equations.

### **References.**

Rubinstein, R.Y. Simulation and the Monte Carlo Method. John Wiley and Sons, New York, 1981.

Papoulis, A. Probability, Random Variables, and Stochastic Processes. McGraw-Hill Book Co, New York, 1965.

Shrieder, Y. The Monte Carlo Method. Pergamon Press, Oxford, 1966.

Carslaw, H. and Jaeger, J. Conduction of Heat in Solids. Oxford University Press, Oxford, 1959.

Figure 1

Monte Carlo Solution with Time =10, Probability =0.1

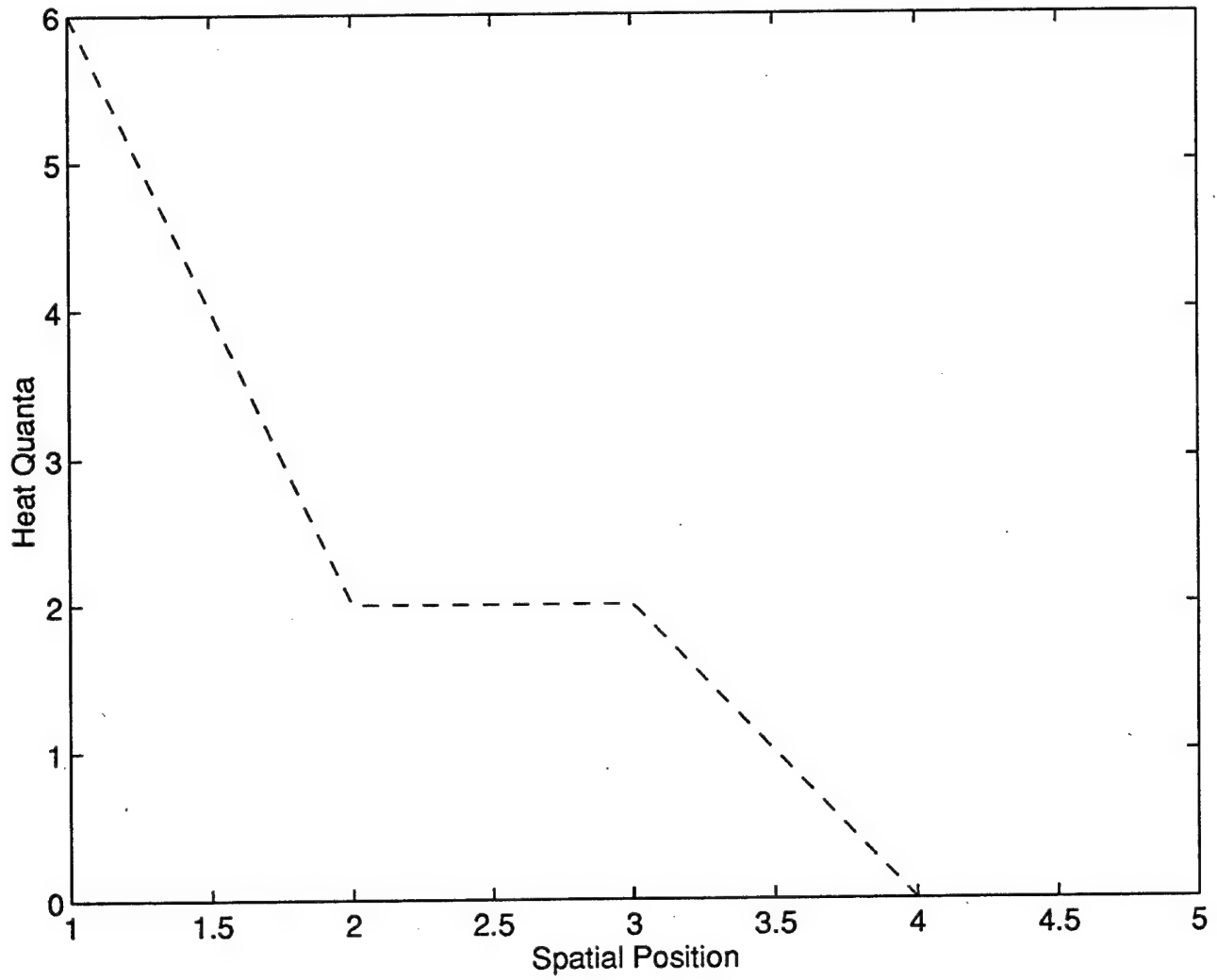


Figure 2

Monte Carlo Solution with Time =25, Probability =0.1

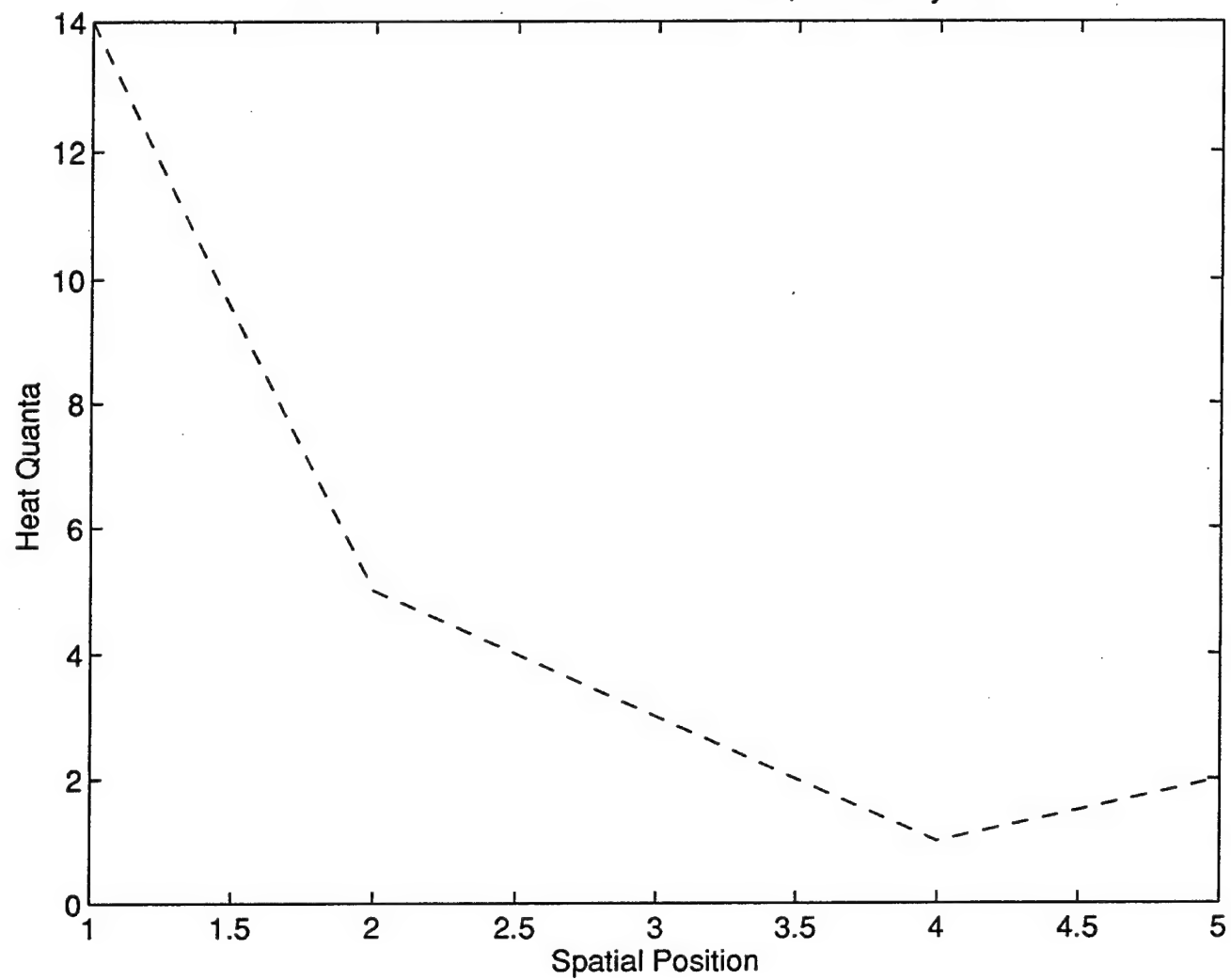


Figure 3

Monte Carlo Solution with Time =50, Probability =0.1

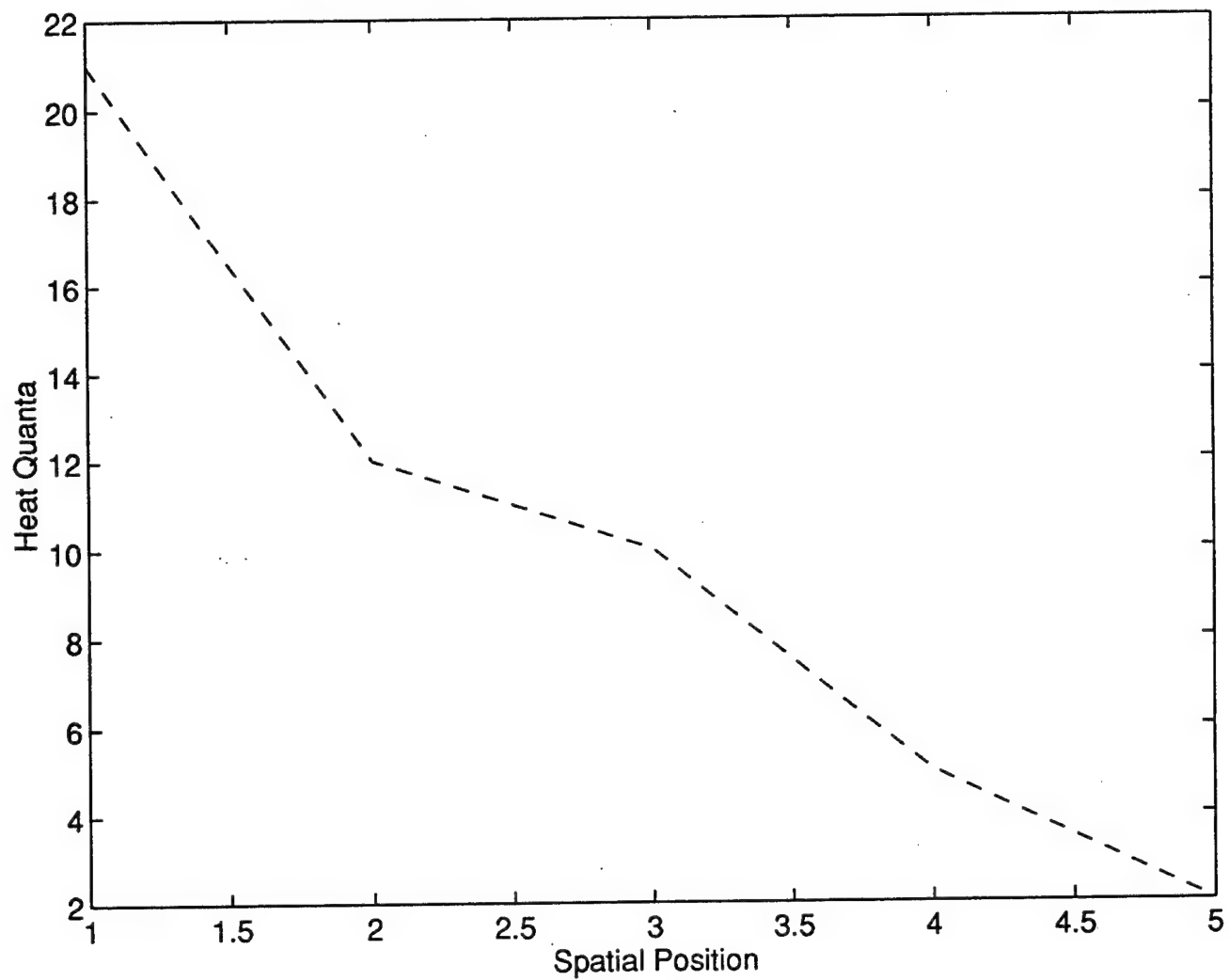
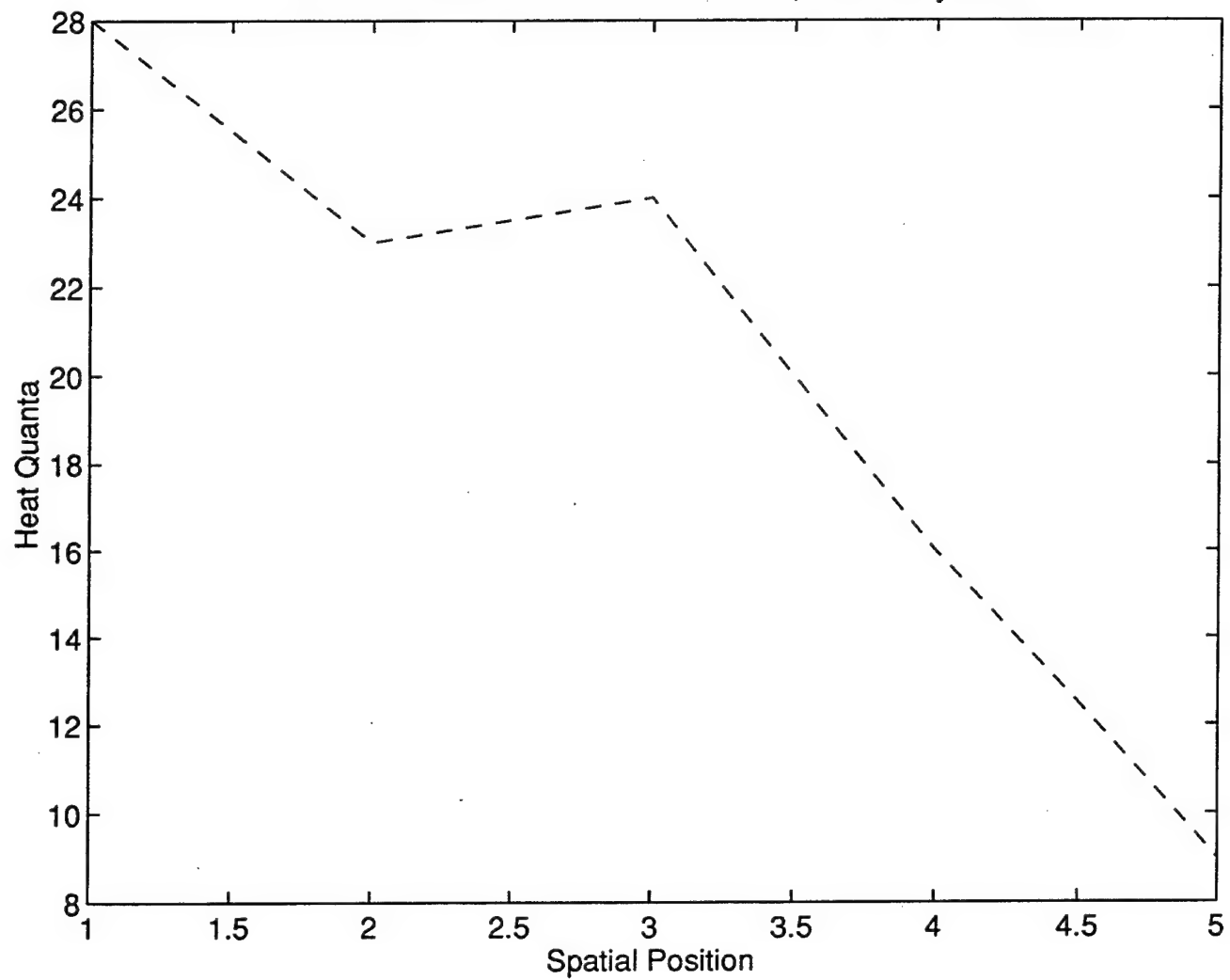


Figure 4

Monte Carlo Solution with Time =100, Probability =0.1



MOOD CHANGE: DOES IT AFFECT  
PSYCHOMOTOR PERFORMANCE?

Thomas L. Dallam  
Doctoral Student  
Department of Psychology

Virginia Polytechnic Institute and State University  
Blacksburg, VA 24061

Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995



## MOOD CHANGE: DOES IT AFFECT PSYCHOMOTOR PERFORMANCE?

Thomas L. Dallam  
Doctoral Student  
Department of Psychology  
Virginia Polytechnic Institute and State University

### Abstract

The purpose of this study was to investigate the effects of mood on psychomotor performance. Subjects were administered the same psychomotor test battery on two different days. Subjects also filled out a mood survey in the morning and afternoon on each of these days. Positive mood was hypothesized to be positively related to psychomotor performance, and negative mood was hypothesized to be negatively related to psychomotor performance. A number of statistical analyses were performed (e.g., hierarchical regression, simple linear regression, MANOVA), but in all, positive and negative mood did not have a significant relationship with psychomotor performance. A number of explanations are offered for these findings.

## MOOD CHANGE: DOES IT AFFECT PSYCHOMOTOR PERFORMANCE?

Thomas L. Dallam

### Introduction

Over the past ten years, the Learning Abilities Measurement Program (LAMP) has been in operation within the Human Resources Manpower and Personnel Division at Armstrong Laboratories. Two initial purposes of this program have been to identify components of the human cognitive system and to assess the generalizability of these components. The ultimate goal of this program is to develop a battery that can successfully be used for aptitude measurement, cognitive task analysis, performance assessment, and training systems for Air Force personnel.

Perceptual-motor abilities are thought to play a large role in performance, especially in tasks such as piloting aircraft and operating naval landing craft and construction equipment. In order to address this facet of performance, the Psychomotor Cognitive Abilities Measurement (PCAM) battery was developed. Based on Fleishman's (1964) psychomotor taxonomy, PCAM tasks were developed to measure hand-eye coordination, reaction time, and the use of joysticks and footpedals.

One concern with test batteries in general and the PCAM specifically, is that non-cognitive variables such as mood may affect individuals' performance. Without measuring mood, this variance is error variance, which attenuates the predictive validity of the test battery. It is the purpose of the present study to examine potential effects of mood on psychomotor performance in order to explain any error variance associated with this test battery. One implication of this study is that, if the hypotheses of this study are supported and it is found that elated mood leads to an increase in performance and depressed mood leads to a decrease in performance, organizations may be able to increase productivity levels by putting their employees in good moods. Organizations may accomplish this by playing uplifting music in the workplace, giving verbal praise, treating employees with respect, etc.

### *Mood*

Although a commonly used term, the definition of mood is often blurred with the construct emotion. Morris (1989) asserts two major differences between mood and emotion: (1) mood is capable of altering responses to a wide variety of objects and events, whereas emotion instigates a relatively limited set of responses, and (2) mood is typically less intense than emotion.

Mood can also be distinguished from affectivity or temperament. The latter represents a trait and is stable, whereas mood represents a state and is transient. Mood has typically been characterized as either positive (happy, enthusiastic, elated, self-assured) or negative (angry, sad, anxious, upset).

Mood has been studied extensively over the past twenty years. Positive mood research has resulted in many robust findings. For example, it has been found to promote self-efficacy (Hill & Ward, 1989; Kavanagh & Bower, 1985), creativity (Isen, Daubman, & Nowicki, 1987; Ross, 1993), risk taking (Isen, Means, Patrick, & Nowicki, 1982), helping behaviors (Isen, 1984), and favorable evaluations of others (Sinclair, 1988) to name but a few. Negative mood research, on the other hand, has been less conclusive. Sometimes it has been found to promote variables like helping behavior or self-efficacy, other times it has been found to decrease them, and still other times it has been found to have no relationship (Isen, 1984). Regardless, both positive and negative mood are presumed to have a great effect on our daily lives.

*Mood and Performance.* Most relevant to the present study is the effects of mood on psychomotor performance. Unfortunately, to the author's knowledge, no research has been conducted on the effects of mood on psychomotor performance. Research has shown, however, that positive mood results in increased physiological arousal and motivation, whereas negative mood produces low arousal and low motivation (Bodenhausen, Sherrard, & Kramer, 1994; Kavanagh, 1987). Given that psychomotor tasks measure reaction time (arousal) and the tasks are arduous (motivation), it would seem that positive mood would facilitate performance and negative mood would hinder performance. Therefore, hypothesis 1 states that positive mood will lead to an increase in psychomotor performance, and hypothesis 2 states that negative mood will lead to a decrease in psychomotor performance.

## METHOD

### *The Larger Study*

This study was conducted as part of a larger study that spanned across eight weeks, five days a week. In addition to the PCAM battery and the mood survey, the larger study consisted of the Cognitive Abilities Measurement (CAM) battery, the driving simulator (Atari videogame where subjects were required to drive a car on the screen), the Vocational Interest for Career

Enhancement Instrument (VOICE - measures an individual's preferences for different types of work-related activities), the Broadbent test, Christal's (1994) Big Five Questionnaire and Self-Description Survey (designed to measure personality traits), and temporal processing task (tasks that required subjects to estimate the passage of time). These other tests are not relevant to the present study and so will not be discussed here. To get a picture of how this larger study was organized, see Table 1 below, which lists the above tests and the time (morning or afternoon) and day in which they were given. This format was the same across all eight weeks of testing.

	Monday	Tuesday	Wednesday	Thursday	Friday
AM	Mood Survey PCAM Battery	Mood Survey CAM Battery	Mood Survey PCAM Battery Voice Survey Broadbent Test	Mood Survey PCAM Battery (shortened)	Mood Survey
PM	PCAM Battery (continued) Mood Survey	Drive Simulator Mood Survey	PCAM Battery (continued) Mood Survey	Big 5 Quest. Self-Description Temp. Process. Mood Survey	Big 5 Quest Self-Description Drive Simulator Mood Survey

Table 1. Tests administered in larger study, ordered by time and day.

### *The Mood and Psychomotor Performance Study*

#### *Subjects*

A total of 161 individuals were recruited from a temporary employment agency to participate in this study, with the requirement that they were similar in demographics to Air Force personnel (i.e., 18-26 years of age, 4:1 male to female ratio). The study spanned across eight weeks, with a different group of subjects participating each week. Because no differences were expected among the eight groups, the groups were combined for data analyses. The sample size varied for each test, as some subjects were not able to complete all of the tests in a given day. For most tests, however, the sample size was at least 145. Subjects were paid eight dollars per hour for their participation.

#### *General Overview*

Subjects filled out a mood scale in the beginning and at the end of each day of testing. On day one, subjects completed the Perceptual-motor Cognitive Abilities Measurement (PCAM) battery. The PCAM battery consisted of 20 different subtests, which were presented to each

subject in random order. Testing lasted seven hours. Subjects then reported back to the testing facility two days later to retake the PCAM battery and mood scales. Again, the individual subtests were presented in random order to each subject, and the testing lasted seven hours.

### *Testing Facility*

The Test Development Center, located at Lackland Air Force Base, consisted of 30 testing stations in a large room. Each station contained a microcomputer system, a standard keyboard for response entry, two joysticks and two foot pedals for psychomotor response entry, and a display monitor. All test materials, including items, response scoring and recording procedures, and feedback presentation procedure, were written in PLATS, a high-level cognitive task authoring system.

### *The Perceptual-Motor Cognitive Abilities Measurement (PCAM) battery*

The perceptual-motor tests were designed to measure abilities that include coordinated movements of two or more limbs, precisely controlled movements in response to dynamic stimuli, speeded movements, and steadiness of arm and hand movements. The tests used in this study were based on a larger taxonomy developed by Fleishman (1964), in which he described eleven psychomotor ability factors. The factors thought to be most related to Air Force tasks were examined in the present study. These factors were control precision, multilimb coordination, response orientation, and rate control. See Table 2 below for a description of these factors and their respective indicators (i.e., subtests).

Factor	Description	Indicator
Control Precision	Fine, highly controlled arm-hand and leg movements. Important in the operation of equipment that requires careful positioning of the feet and hands. It is especially critical when such movements require speed with precision.	CP1 - Rotary Pursuit (follow moving target using joystick) CP2 - Tracking Targeting Task (track and shoot moving helicopter) CP3 - Joystick Swing Task (move sight from strip to another) CP4S - Small Grid (click on square before it moves to new location) CP4T - Tiny Grid (same as above, only smaller squares)
Multilimb Coordination	Simultaneous coordination of two or more limbs especially when operating devices with several controls.	ML1 - Center the Ball (keep green ball centered on yellow crosshairs using joystick and footpedals) ML2 - Complex Coordination Task (line up green lights with red lights using joystick and foot pedals) ML3A - Two Arm Pursuit (tests ability to use dual joysticks) ML3B - Star Draw (tests ability to use two joysticks and foot pedals) ML4 - Pop the Balloons (track & pop balloons using foot pedals & joystick)
Response Orientation	Rapid directional discrimination and orientation of movement patterns. This factor enables the selection of the correct response given a stimulus independent of either precision or coordination.	RO1 - Lights & Sounds (react quickly to lights and sounds using joystick and foot pedals) RO2 - Red to Green Orientation (respond with # pad keys according to a set of given rules) RO3 - Directional Control (respond with joystick and buttons according to a set of given rules) RO4 - Touchscreen Task (touch certain areas of screen according to set of given rules)
Rate Control	Making continual corrections to a motor response to keep in synchrony with changes in the speed and direction of a continuously moving target.	RC1 - Arc Pursuit (keep red target between 2 yellow lines) RC2A - Motor Judgment Task (keep green bar out of red sections) RC2B - Single Wheel Task (keep green bar out of gray sections) RC3 - Circular Pursuit (Keep yellow cursor on moving red target) RC4 - Lane Tracking (keep car on center line despite wind gusts)

Table 2. Summary of psychomotor domains and their respective tests.

### *Mood Survey*

The mood survey used in this study was developed by Christal (1994). The survey was based on Watson, Clark, and Tellegen's (1988) Positive and Negative Affect Schedule (PANAS). Seven mood categories were measured. These categories are defined in Table 3 below. In addition, two higher order categories were formed; one composed of all of the negative mood descriptors (i.e., the descriptors underlying the fear, sadness, guilt, and hostility domains), and the other composed of all the positive mood descriptors (i.e., the descriptors underlying the joviality, self-assured, and attentiveness domains). Thus, positive mood was operationally defined as the average of all of the positive mood descriptors, and negative mood was operationally defined as the average of all of the negative mood descriptors. The positive and negative mood categories are also presented in Table 3 below.

Category	Items Included
1. Fear	Scared; Nervous; Jittery; Afraid; Frightened; Shaky
2. Sadness	Sad; Blue; Downhearted; Alone; Lonely
3. Guilt	Guilty; Ashamed; Blameworthy; Dissatisfied with yourself; Angry at yourself; Disgusted with yourself
4. Hostility	Hostile; Irritable; Angry; Scornful; Disgusted
5. Joviality	Enthusiastic; Excited; Happy; Delighted; Cheerful; Joyful; Lively; Energetic
6. Self-Assurance	Strong; Proud; Confident; Bold; Daring; Fearless
7. Attentiveness	Alert; Determined; Attentive; Concentrating
8. Negative	Distressed, Upset, Guilty, Scared, Hostile, Irritable, Ashamed, Nervous, Jittery, Afraid, Frightened, Shaky, Sad, Blue, Downhearted, Alone, Lonely, Blameworthy, Dissatisfied, Angry, Angry at yourself, Disgusted, Disgusted with yourself, Scornful
9. Positive	Interested, Strong, Enthusiastic, Proud, Alert, Inspired, Determined, Attentive, Active, Excited, Happy, Delighted, Cheerful, Joyful, Lively, Energetic, Confident, Bold, Daring, Fearless, Focused

**Table 3.** Mood scale factors and the indicators that make up each factor.

Subjects responded to each of these descriptors by indicating the extent to which each item described how they currently felt. The scale was presented to them on a computer screen with the descriptor in a box at the top and an arched gauge below. The gauge consisted of five different verbal anchors: slightly or not at all, a little bit, moderately, quite a bit, and extremely. Subjects recorded their response by moving the mouse cursor along one of 44 different hidden points on the gauge and clicking the mouse button. The hidden points ranged from 1 to 44, with a higher number representing a higher degree to which the subject felt that particular descriptor.

### *Dependent Measures*

Depending upon the nature of the test, either accuracy or time served as the dependent measures for each of the psychomotor tests. Accuracy and time were measured in various ways depending upon the type of psychomotor test. See Table 4 below for a listing of the ways these dependent variables were measured.

Psychomotor Test	Accuracy Measure	Reaction Time Measure
CP1	Avg distance from ball	Avg time taken to complete set Avg time taken to complete set Avg time taken to complete set
CP2	Number of helicopters shot	
CP3		
CP4S		
CP4T		Average time Avg time taken to pop balloons
ML1	Avg distance from center	
ML2	Avg number completed	
ML3A	Avg distance	
ML3B		Avg time to respond correctly Average time to respond correctly Average response time
ML4	Number of balloons popped	
RO1	Number correct	
RO2		
RO3		
RO4		
RC1	Average distance from target	
RC2A	Passes minus # of errors	
RC2B	Avg # of errors	
RC3	Avg distance from target	
RC4	Avg distance from center lane	

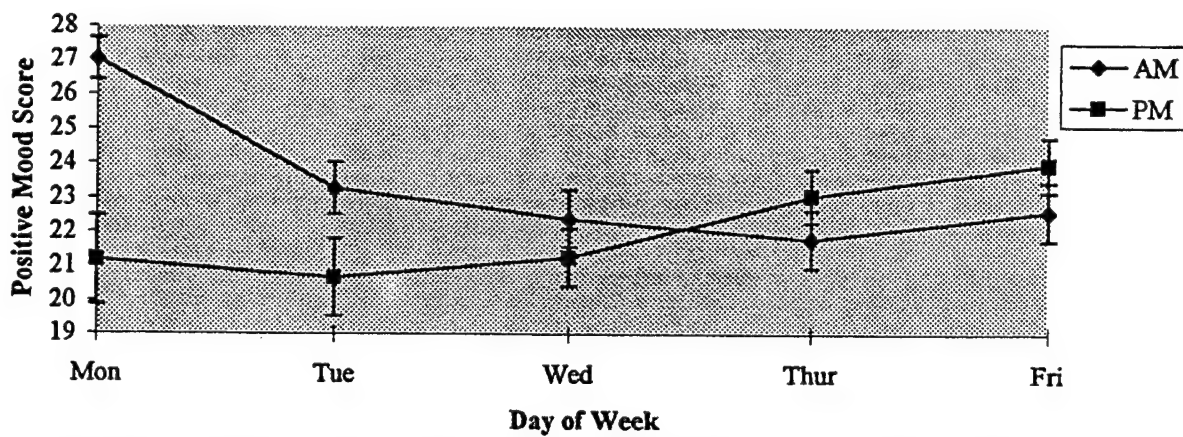
Table 4. Dependent variables used for each of the tests in the psychomotor test battery.

## RESULTS

### *Mood Changes*

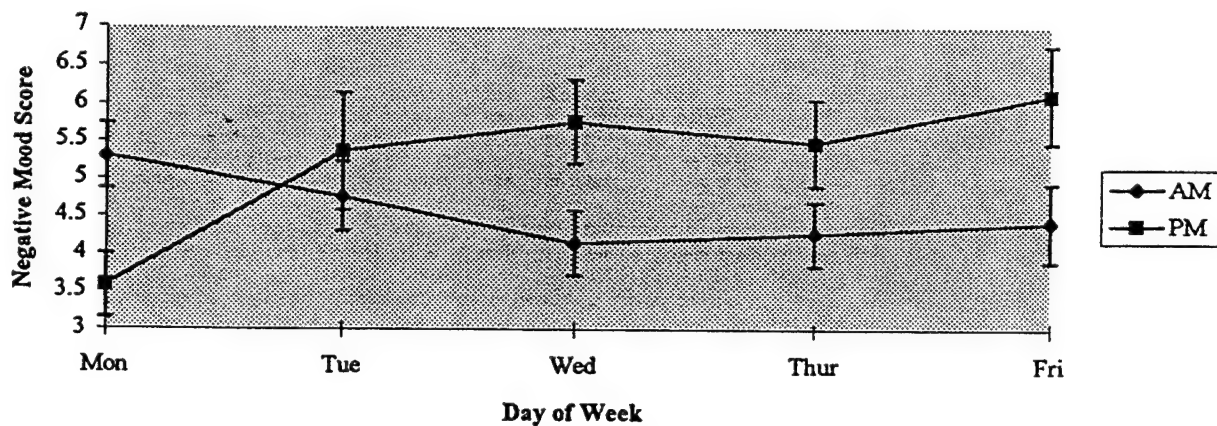
One interesting question posed by this study was whether subjects' mood would change in response to the testing sessions. Although no formal hypotheses were submitted, one might expect, due to the challenging and exhausting nature of the tests, that positive mood would decrease and negative mood would increase as the day wore on. Some support for the positive mood notion was found. In looking at Figure 1 below, one can see that positive mood did decrease significantly from morning to afternoon on day one. This decrease was not found on day two, but that seems to be due to a floor effect, as subjects started out with a much lower positive mood on day two than on day one. Thus, the testing effects may have carried over to day two.





**Figure 1.** Changes in positive mood from morning to afternoon on each day of testing.

Negative mood results were much more perplexing. Subjects' negative mood actually decreased from morning to afternoon on day one, and increased from morning to afternoon on day two (see Figure 2 below).



**Figure 2** Changes in negative mood from morning to afternoon for each day of testing.

### Mood and Psychomotor Performance

#### *Mood Estimation Models*

Although the testing session lasted throughout the day (seven hours), mood was only measured at the beginning and end of the day. It is quite possible that subjects' moods varied throughout the day in the time between the two mood measures. Thus, an estimate must be made

of subjects' mood at the time of each psychomotor test. This estimation was calculated in two different ways, which are presented below along with the ensuing results.

*Averaged Mood Estimation.* This measure was computed by averaging subjects' scores on the morning mood scale and their scores on the afternoon mood scale for day two. This scores served as a predictor for performance on each psychomotor test on day two. In order to determine whether this mood score contributed unique variance to the prediction of day two performance, a hierarchical regression was performed by entering psychomotor performance on day one first into the equation and mood score second. Positive and negative mood scores were analyzed separately for each of the twenty different psychomotor test scores, producing a total of forty different hierarchical regression analyses. Only four of these regressions resulted in a significant change in  $R^2$ , two for positive mood and two for negative mood. The significant regressions are presented in Table 5 below.

Criterion	Mood	N	b	SE b	$\beta$	$\Delta R^2$
CP1	Positive	121	-4.54	2.15	-.13	.02
ML3A	Positive	120	-3.48	1.51	-.11	.01
RC1	Negative	119	2.08	.85	.16	.03
RO3	Negative	116	12.92	6.5	.12	.01

Table 5. Hierarchical regression results for averaged mood estimation scores (significant findings only).

Given the high number of univariate regressions conducted, the significant findings above are likely due to alpha inflation error. That is, with 40 regressions one would expect two (5%) of them to be significant at the .05 level. Thus, according to the averaged mood estimation technique, mood does not seem to add incremental validity over past psychomotor performance.

*Linear Mood Estimation.* This measure assumes that mood change is linear. That is, mood state monotonically increases or decreases throughout the day, depending upon what the mood state is at the end of the day. This measure was computed by weighting both the morning and afternoon score in terms of the exact time in which they were taken. For example, in computing subjects' positive mood while taking the CP1 psychomotor test on day one, the following formula would be applied:

$$[(MSPM_{1TM} - CP1_{1TM}) / (MSPM_{1TM} - MSAM_{1TM})] * POSAM_{1J} + [(MSAM_{1TM} - CP1_{1TM}) / (MSAM_{1TM} - MSPM_{1TM})] * POSPM_{1J}$$

where the subscript <sub>1</sub> indicates day one testing; MSPM1TM = the time at which the afternoon mood survey was taken; CP1<sub>1</sub>TM = the time at which the CP1 test was taken; MSAM<sub>1</sub>TM = the time at which the morning mood survey was taken; POSAM<sub>1</sub> = subjects positive mood score on the morning mood survey; and POSPM<sub>1</sub> = subjects' positive mood score on the afternoon mood survey.

Hierarchical regression analyses were again used to determine the incremental validity of these mood scores. Positive and negative mood scores were analyzed separately for each of the psychomotor tests, producing a total of 40 different hierarchical regression analyses. In only a handful of these regressions was mood found to contribute unique variance over past psychomotor performance. The significant regressions are presented in Table 6 below.

Criterion	Mood	N	b	SE b	$\beta$	$\Delta R^2$
CP3	Positive	112	-2.92	1.20	-.17	.03
ML3A	Positive	104	-3.41	1.52	-.11	.01
RC1	Negative	103	3.02	.93	.23	.05
RC4	Negative	113	.99	.50	.14	.02
RO3	Negative	100	22.31	7.91	.18	.03

**Table 6.** Hierarchical regression results for linear mood estimation scores (significant findings only).

As with the average mood estimation technique, the linear mood estimation technique does not provide support for both hypotheses. That is, as a whole, positive mood does not lead to an increase in psychomotor performance and negative mood does not lead to a decrease in psychomotor performance. In fact, for the few significant findings, the pattern was in the opposite direction. Positive mood had a negative beta weight and negative mood had a positive beta weight (see Table 6 above).

*Juxtaposed Tests.* It is not unreasonable to question the validity of the two mood estimation techniques presented above, especially given the relatively few significant findings. Perhaps, it is more valid to examine only those psychomotor tests that were taken either immediately after the administration of the mood scale or before the administration of the mood scale. Certainly, it requires less of an inference to establish that subjects' mood scores represented the mood they were in during the administration of that one psychomotor test.

Thus, a series of univariate regressions were computed; one using scores from the first psychomotor test taken after the morning mood survey on day one, and one using scores from the last psychomotor test taken immediately before the afternoon mood survey on day one. The same procedure was used for day two testing, producing four different regression analyses. Within each regression procedure, positive mood and negative mood scores from the adjacent mood survey were entered as predictors. Before presenting these results, however, recall that the order in which subjects were administered the psychomotor tests was randomized, meaning that not everyone took the same test immediately after or before filling out the mood survey. Therefore, scores on these tests were standardized (i.e., z-scores) in order to put them on the same metric. The results are summarized in the Tables 7 and 8 below.

Day One						
Criterion	Predictor	b	SE b	T	R <sup>2</sup>	N
First Test	Pos. Mood	.006	.012	0.467	.001	157
	Neg. Mood	.002	.017	.136		
Last Test	Pos. Mood	.022	.014	1.514	.08	50
	Neg. Mood	.067	.046	1.469		

Table 7. Regression results using adjacent tests on day one as criteria.

Day Two						
Criterion	Predictor	b	SE b	T	R <sup>2</sup>	N
First Test	Pos. Mood	-.004	.008	-0.508	.04	146
	Neg. Mood	.042	.018	2.356*		
Last Test	Pos. Mood	.003	.008	.326	.001	130
	Neg. Mood	.004	.013	.333		

Table 8. Regression results using adjacent test on day two as criteria.

### *Multivariate Analysis.*

In order to examine whether mood had an effect on psychomotor performance as a whole, a series of MANOVAs were conducted. The 20 psychomotor tests were submitted as the dependent variables. Mood was again determined by using the averaged mood estimation. Thus, four MANOVAs were examined: Positive mood on day one, positive mood on day two, negative mood on day one, and negative mood on day two.

	Positive Mood		Negative Mood	
	Day One	Day Two	Day One	Day Two
N	47	113	47	113
F	.46	.19	.71	.22

Table 9. MANOVA results for both positive and negative mood using the averaged mood estimation for day one and day two.

### *Factor Scores.*

A factor analysis was conducted on the 20 psychomotor tests taken by the subjects. This analysis was performed in order to use factor scores as dependent variables which, given the high number of tests, provides a more parsimonious analysis of the effects of mood on psychomotor performance. The principle axis procedure with a varimax rotation was conducted separately for day one psychomotor performance and day two psychomotor performance. Based on the scree plot and eigenvalues, this analysis produced four factors for each set of data. The four factors for day one accounted for 65 percent of the common variance, and the four factors for day two accounted for 67 percent of the common variance. The rotated factor matrix patterns for day one and day two are presented below.

Day One				
	Factor 1	Factor 2	Factor 3	Factor 4
RO3	.68			
ML1	.67			
ML2	-.57			
ML3A	.57			
RO4	.55	.44		
ML3B	.48	.42		
RO2	.47	.38		
RO1	.28	.25	.18	.22
RC3		.87		
RC1		.59	.52	
RC2A		-.53		
CP2	-.44	-.48	-.44	
RC2B		.43		
CP1			.92	
RC4	.39		.48	
ML4		-.38	-.44	
CP4T				.98
CP4S				.50
CP3	.35			.42
RO4	-.25			.37

Table 11. Rotated factor matrix pattern of PCAM for day one.

Day Two				
	Factor 1	Factor 2	Factor 3	Factor 4
RC3	.76			
RC2	.72			
ML3B	.66			
ML4	-.61		-.40	
ML2	-.53			
RO3	.52			
RO4	.33	.29		
CP1		.87		
RC4		.61	.45	
RC1	.39	.58	.37	
ML3	.44	.51	.39	
RO1		.31		
ML1		.46	.65	
RC2			-.57	
CP2	-.49		-.53	
RO2			.52	
CP4T				.92
CP4S				.75
CP3				.45
RO4				.34

Table 12. Rotated factor matrix pattern of PCAM for day two.

These factors were regressed on positive and negative mood. Again, mood score was computed through the averaged mood estimation technique. The results are presented in Tables 13 and 14 below.

Day One (N=47)					
Criterion	Predictor	b	SE b	T	R <sup>2</sup>
Factor 1	Pos. Mood	.007	.021	0.333	.02
	Neg. Mood	.045	.046	.991	
Factor 2	Pos. Mood	-.010	.018	-.563	.03
	Neg. Mood	.040	.041	.973	
Factor 3	Pos. Mood	-.006	.021	-.287	.01
	Neg. Mood	.018	.046	.389	
Factor 4	Pos. Mood	-.010	.021	-.483	.01
	Neg. Mood	.026	.045	.576	

Table 13. Regression results for day one, using factor scores as dependent measures and averaged mood as predictors.

Day Two (N = 113)					
Criterion	Predictor	b	SE b	T	R <sup>2</sup>
Factor 1	Pos. Mood	.008	.010	0.838	.02
	Neg. Mood	.024	.020	1.195	
Factor 2	Pos. Mood	-.006	.010	-.545	.03
	Neg. Mood	.034	.021	1.66	
Factor 3	Pos. Mood	-.009	.010	-.883	.01
	Neg. Mood	-.017	.021	-.794	
Factor 4	Pos. Mood	-.009	.011	-.864	.01
	Neg. Mood	.001	.022	.051	

Table 14. Regression results for day two, using factor scores as dependent measures and averaged mood as predictors.

Again, both positive and negative mood failed to predict psychomotor performance.

## DISCUSSION

The purpose of this study was to investigate the effects of positive and negative mood on psychomotor performance. It was hypothesized that positive mood would facilitate psychomotor performance and negative mood would hinder it. The results did not support these hypotheses, however. When using the averaged mood estimation (averaging the morning and afternoon mood scores on day two), hierarchical regression analyses showed that when taking past psychomotor

performance into account, positive and negative mood each contributed unique variance on only two of the 20 psychomotor tests (CP1 and ML3A for positive mood; RC1 and RO3 for negative mood). Essentially the same results were found when using the linear mood estimation (using times at which tests were taken as weights), as positive mood contributed unique variance on only two of the psychomotor tests (CP3 and ML3A) and negative mood on only three (RC1, RC4, and RO3). Two possible explanations exist for these significant findings. One is that there may be something particular about these psychomotor tests that invites positive or negative mood as a predictor. Perhaps these tests were harder than the others, requiring more effort which has been found to be affected by mood (Kavanagh, 1987). The more likely explanation, however, is that these significant findings were due to chance. As any researcher will attest, alpha inflation error becomes a problem when a high number of tests are conducted. In this case, a total of 40 hierarchical regressions were computed, of which two would be expected to be significant by chance.

Other statistical analyses bore the same results. In order to assess the relationship between mood and psychomotor performance as a whole, a multivariate analysis of variance was conducted with all 20 psychomotor tests serving as dependent variables. Neither positive nor negative mood significantly predicted psychomotor performance on day one or day two. Another method for looking at psychomotor performance as a whole was to factor analyze the psychomotor tests and submit the resulting factor scores as dependent variables in a regression analysis. This factor analysis produced four factors for both days of testing. With positive and negative mood as predictors, regressions were performed for each day of testing. Again, no significant results were found.

As any researcher will attest, null results are the bane of our profession because it is too difficult to determine the exact reason for this lack of significant differences. The present study is no exception here. It is posited that three different explanations can account for these results.

One concern with this study is that the variance and the means for the negative mood items were quite low, which can attenuate regression coefficients. With a range of 44, the mean for negative mood was 3.81 and the standard deviation was 4.37 across all days of testing. Clearly, people were either not in a bad mood, or they were not reporting it. Compare these numbers with other studies using similar mood scales and you will find they are lower. For



example, Watson, Clark, and Tellegen (1988) reported a mean of 14.8 (out of a possible 50) and a standard deviation of 5.4. Kennedy-Moore, Greenberg, Newman, and Stone (1992) reported a mean of 12.65 and a standard deviation of 3.67. Thus, the lack of variance does not seem to be due the scale itself, but rather to the subjects. It is possible that individuals did not feel comfortable in stating their negative feelings. They may have wanted to present themselves in a more positive light (i.e., social desirability bias), or maybe they were just not in tune with their emotions. Regardless, this low variance may have played a part in the null results.

A second possible reason for these null results centers around the concept of self-efficacy (i.e., confidence that a person can perform a particular task successfully). Some researchers theorize that self-efficacy is a full mediator of mood and performance (K. Williams, personal communication, 1995). If this theory is indeed true, then the relationship between mood and performance would go to zero if self-efficacy is not measured. Since self-efficacy was not measured in the present study, this explanation remains to be seen.

A third possible reason for the lack of a mood relationship is simply that mood really does not affect psychomotor performance. Clearly, before this assertion can be stated with confidence, more studies need to be conducted addressing some of the concerns mentioned here. Future studies should ensure there is enough variance in positive and negative mood. A good practice would be to actually induce the moods. Many mood induction procedures exist, with music, videotapes, and false performance feedback serving as some of the most valid ways of manipulating mood (Gerrards-Hesse, Spies, & Hesse, 1994). Also, a measure self-efficacy should be included in the study. Mediated regression analysis should then be conducted to determine if self-efficacy is either a partial or full mediator of the mood-performance relationship. Until actions such as these are taken, the relationship between mood and psychomotor performance will remain unclear.

## References

- Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. Special Issue: Affect in social judgments and cognition. *European Journal of Social Psychology*, 24, 45-62.
- Christal, R. E. (1994). Non-cognitive research involving systems of testing and learning (ASD) (Contract No. F33615-91-D-0010). Armstrong Laboratory, Brooks AFB, San Antonio, TX.
- Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, 85, 55-78.
- Hill, R. P., & Ward, J. C. (1989). Mood manipulation in marketing research: An examination of potential confounding effects. *Journal of Marketing Research*, 26, 97-104.
- Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition* (pp. 174-236). Hillsdale, NJ: Erlbaum.
- Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, 52, 1122-1131.
- Isen, A. M., Means, B., Patrick, R., & Nowicki, G. P. (1982). Some factors influencing decision-making strategy and risk taking. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition: The 17th Annual Carnegie Symposium on Cognition* (pp. 243-261). Hillsdale, NJ: Erlbaum.
- Kavanagh, D. J. (1987). Mood, persistence, and success. *Australian Journal of Psychology*, 39, 307-318.
- Kennedy-Moore, E., Greenberg, M. A., Newman, G., and Stone, A. A. (1992). The relationship between daily events and mood: The mood measure may matter. *Motivation and Emotion*, 16, 143-155.
- Kyllonen, P. C. (1994). Cognitive abilities testing: An agenda for the 1990's. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel Selection and Classification*. Hillsdale, NJ: Lawrence Erlbaum.
- Sinclair, R. C. (1988). Mood, categorization breadth, and performance appraisal: The effects of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes*, 42, 22-46.

Watson, D., & Clark L. A. (1992). On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. *Journal of Personality*, 60, 441-476.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.

**Eric Fetterhoff's Report not available at time of publication.**

ELECTROPHYSIOLOGICAL, BEHAVIORAL, AND SUBJECTIVE INDEXES  
OF WORKLOAD WHEN PERFORMING MULTIPLE TASKS:  
MANIPULATIONS OF TASK DIFFICULTY AND TRAINING

Lisa R. Fournier  
Graduate Student  
Department of Psychology

University of Illinois  
603 E. Daniel  
Champaign, IL 61821

Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, Washington, DC

and

Armstrong Laboratory

August 1995

ELECTROPHYSIOLOGICAL, BEHAVIORAL, AND SUBJECTIVE INDEXES  
OF WORKLOAD WHEN PERFORMING MULTIPLE TASKS:  
MANIPULATIONS OF TASK DIFFICULTY AND TRAINING

Lisa R. Fournier  
Graduate Student  
Department of Psychology  
University of Illinois

Abstract

This study examines the effectiveness of using brain activity such as alpha rhythm, event-related desynchronization (ERD) as a measure of mental workload in a "real world" task. Subjects performed multiple tasks that were similar to those encountered by a pilot inside of an airplane cockpit. These tasks were interactive and included monitoring the path of the airplane (tracking), verbal communications, gauge display changes, and light display changes. We were interested in whether the alpha ERD can successfully measure workload as well as discriminate between novice and expert performance obtained through training. We expect alpha ERD duration to increase and alpha amplitude to decrease under conditions of high mental workload. Also, we expect differences in alpha ERD duration and amplitude across the various levels of workload difficulty (low, medium, and high) to be reduced after training. Additional measures, including heart rate, respiration, behavioral performance, and subjective workload measures were recorded to determine how well alpha ERD duration and amplitude correlates with these workload measures. If alpha ERD proves to be a successful index of workload, we can use this measure to determine what characteristics in operator displays (e. g., cockpit displays) increase or decrease operator workload.

# ELECTROPHYSIOLOGICAL, BEHAVIORAL, AND SUBJECTIVE INDEXES OF WORKLOAD WHEN PERFORMING MULTIPLE TASKS: MANIPULATIONS OF TASK DIFFICULTY AND TRAINING

Lisa R. Fournier

The main goal of the present study was to examine the feasibility of employing brain electrophysiological measures to monitor changes in mental workload. We were interested in measuring brain event related changes because behavioral and subjective measures are not always reliable indexes of workload (e. g., Yeh & Wickens, 1988). In addition, measures such as heart rate and respiration do not always allow us to speculate on what type of mental processes are in high demand.

We were specifically interested in the event related amplitude changes or desynchronization (ERD) of the alpha rhythm (Pfurtscheller & Aranibar, 1977; Pfurtscheller & Klimesch, 1989). The usual frequency range for alpha is between 8-12 Hz. In general, alpha rhythm amplitude tends to decrease when a person is alert or engaged in a task and the amplitude tends to increase when a person is in a relaxed state, with eyes closed. The alpha ERD is a topographically localized decrease in alpha amplitude of short duration (phasic) and has been shown to occur during visual stimulation, auditory stimulation, voluntary movement, and cognitive activity. Figure 1 contains an example of an ERD in the alpha band. Attenuation of the alpha rhythm has been shown to be maximal at the occipital electrode sites for visual-spatial tasks (Pfurtscheller & Aranibar, 1977) and to be maximal at centro-parietal sites when movement is required (Pfurtscheller & Aranibar, 1979). In

addition, attenuation of the alpha rhythm has been shown to be maximal over auditory cortex for auditory discrimination tasks (Kaufman, Curtis, Wang, & Williamson, 1991). Furthermore, alpha rhythm ERD amplitude and duration has been shown to vary with task difficulty or cognitive workload. Increases in cognitive workload tends to increase the relative alpha desynchronization amplitude and increase the duration of alpha desynchronization (van Winsum, Sergeant, & Geuze, 1984). Fast event rates (high action demands) within a task has also led to smaller relative desynchronization and shorter ERD than tasks with slow event rates (Geuze & van Winsum, 1987; van Winsum et al., 1984). This latter finding suggests that when events are difficult to predict, alpha desynchronization increases. However, it has been shown that increasing the levels of cognitive load causes the amount of ERD and the duration of the ERD to increase independent of event rate (van Winsum et al., 1984).

Alpha ERD has been successfully used to measure workload in single task paradigms, and these paradigms have typically required the subject to perform simple tasks such as judging whether a letter occurs within a memory set (Geuze & van Winsum, 1987; van Winsum et al., 1984; Kaufman et al., 1991). We were interested in determining whether the alpha ERD could successfully measure workload while a subject is engaged in multiple tasks. In addition, we were interested in whether the alpha ERD can be used to discriminate novice and expert performance obtained through training. Moreover, we wanted to determine whether alpha ERD can successfully index workload and learning in real world



interactive tasks. If alpha ERD proves to be a successful index of workload, we can use this measure to determine what characteristics in operator displays (e. g., cockpit displays) increase or decrease operator workload.

In the present study, subjects performed multiple tasks that were similar to those encountered by a pilot inside of an airplane cockpit. These tasks were interactive and included monitoring a simulated flight path (tracking), verbal communications, gauge display changes, and light display changes. We were mainly interested in measuring the effects of workload (varied by tracking difficulty and event rate of gauge and light changes) on the communication task, which required subjects to classify communications as relevant or irrelevant and carry out instructions given in the relevant messages. We were also interested in how training under the different workload conditions influenced workload demands. We expect that alpha ERD will increase with increases in mental workload. We also expect alpha ERD to decrease across all levels of workload difficulty with training.

In addition to recording brain electrophysiology, other measures including heart rate and respiration, behavioral performance (reaction time and accuracy), and subjective ratings of task difficulty were used to index workload. We will examine how these measures vary with workload and training, and how they correlate with alpha ERD results.

## Method

### Subjects.

Four males and six females ages 18 - 26 participated as paid volunteers. All were right-handed and had normal or corrected-to-normal vision (questionnaire).

### Electrophysiological Recording.

Ag/AgCl electrodes were used. The impedance for EEG and EOG electrodes was less than 5 Kohm and EKG electrodes were less than 10 Kohm.

EKG and Respiration. Heart activity was recorded by electrodes placed on the skin above the sternum and fifth intercostal space of the left ribcage. The ground electrode was located above the fifth intercostal space of the right ribcage. Respiration was recorded using a Resptrace system with elastic transducer bands on the chest and abdomen. Respiration amplitude was calculated for each subject prior to data collection. Heart activity and respiration were sampled continuously at 1000 Hz. Signals were recorded with Grass P511 amplifiers and data were filtered at 10-100 Hz.

### Electroencephalogram (EEG) and Electrooculogram (EOG).

Neuroscan software, amplifiers, and ElectroCap were used to record EEG over 58 scalp sites (consistent with the international 10/20 system, Jasper, 1958). Each scalp electrode was referenced to linked mastoids. Eye blinks were recorded from electrodes placed above and below the right eye to monitor vertical eye movements and from electrodes placed on the outer canthus of the left and right eye to monitor horizontal eye

movements. EOG and EEG signals were recorded with Neuroscan amplifiers (model 5083) with a 60 Hz notch filter. EEG and EOG were sampled continuously at 200 Hz. EEG and EOG signals were filtered at 0.1-30 Hz. Eye-movement artifacts were corrected using Neuroscan's eye-movement correction procedure.

#### Task Apparatus and Stimuli.

Subjects were seated at a table in a sound proof, electrically shielded chamber. A joystick was placed to the right of the subject and a computer mouse and pad was placed to the left of the subject. Subjects controlled the joystick with their right hand and controlled the computer mouse with their left hand. Visual displays were presented on a computer monitor centered 38 inches from the subject. In addition, audio messages were presented over a speaker (Realistic model 32-2040) centered in front of the computer monitor.

Subjects were presented with visual monitoring tasks, an auditory task, and a tracking task from the Multi-Attribute Task Battery (MATB) developed by J. R. Comstock and R. J. Arnegard (1992). MATB was designed to simulate displays encountered in an airplane cockpit. Figure 2 shows the MATB visual displays for each task and these tasks are described below.

Light Detection Task. Subjects were to detect the offset of a green light and the onset of a red light occurring in the upper left (F5) and upper right (F6) of the system monitoring display, respectively. Normally, the green light appeared in the upper left box and the upper right box was black. The boxes subtended approximately 1.2 deg of visual angle.

Subjects indicated that they detected the offset of the green light or the onset of the red light by moving a mouse to the box corresponding to the light fault detected, and then clicking the mouse on this box. This action caused the lights to return to their normal state (either green light on or red light off). The light faults (green off, red on) had a maximum duration of 15 sec.

Gauge Monitoring. Subjects monitored 4 different gauges (2.4 deg visual angle each) and determined whether a yellow pointer fluctuated outside of a specified range. The four gauges (F1, F2, F3, and F4) were located in a horizontal row just below the light boxes. Normally, each yellow pointer independently fluctuated from one unit below to one unit above the center line (see Figure 2). A gauge fault occurred if one of the yellow pointers fluctuated more than one unit above or below the center line. Subjects indicated the detection of a gauge fault by moving the mouse anywhere on the faulty gauge and clicking the mouse. After clicking the mouse on the faulty gauge, a yellow bar (.12 deg visual angle) appeared at the bottom of the gauge, and the fault was corrected. A gauge fault had a maximum duration of 15 sec.

Tracking. The goal of this task was to keep the airplane symbol, represented by a green circle (.6 deg visual angle), in the center rectangular area (1.8 deg visual angle). (See upper right panel in Figure 2). The disturbance function for tracking consisted of a sum of nonharmonic sine waves. The frequency of the disturbance function varied for low, medium, and high tracking according to the MATB program specifications. Corrective movements of the airplane symbol

was controlled by moving the joystick (first order control). When the cross-hatches of the circle and the cross-hatches in the center rectangle were aligned, this represented optimal tracking performance.

Communication. Audio messages were 6 sec in duration and consisted of a 6 digit call sign followed by a command. An example of a message is as follows, "NGT504, NGT504, set first communication to 119.7". The relevant target callsign for each subject was "NGT504". Any other call sign (i. e., "NRK362", "NLS217", "NAL478", "NDL183") was irrelevant (noise) and was to be ignored. The command indicated which channel on the computer screen (lower left in Figure 2) needed to have its frequency adjusted. Four possible channels required adjustments: NAV1, NAV2, COM1, and COM2. These were verbally indicated in the message as "first navigation", "second navigation", "first communication", and "second communication", respectively. A channel was selected by moving the mouse on the channel name (e. g., NAV1) and clicking the mouse. The number to the right of the channel name was the corresponding frequency of that channel. The frequency of the selected channel was changed by clicking the mouse on the arrow icons located to the right of the frequency numbers (see Figure 2). Selecting the right-most arrow and clicking the mouse caused the frequency to increase (increments of 0.2 units) and clicking the mouse on the left-most arrow caused the frequency to decrease (decrements of 0.2 units). The navigation frequencies ranged from 108.1 to 117.9 and the communication frequencies ranged from 118.1 to 135.9. After the frequency changes were made, the mouse then clicked on the "Enter" box

(1.2 deg vis angle) located below the channel names. The "Enter" box borders brightened to indicate that the response was recorded. Text characters (channel name, frequency, and arrows) measured .36 deg visual angle.

Subjective workload rating scale. The NASA Task Load index (NASA TLX) developed by Hart and Staveland (1988) was used to obtain a self reported assessment of the subjects workload during the multi-task conditions (see procedure description of conditions). This 10 point multi-dimensional scale is shown in Figure 3. There were a total of six subscales. Subjects rated their exertion from low (1) to high (10) in the following 5 subscales: Mental demand, physical demand, performance, temporal demand, effort, and frustration. Subjects completed the NASA TLX for the 3 levels multi-task conditions (low, med, and high), and each of the task events within each of the multi-task conditions (light detection, gauge monitoring, tracking, and communication). Thus, a total of 5 subscales for each multi-task condition (low, med, and high) was obtained.

#### Procedure.

Practice. At the beginning of the experiment, subjects practiced controlling the mouse with their left hand. They moved the mouse and clicked on the different task locations within the MATB displays for approximately 9 min.

Experimental Sessions. Each subject was run through four different task conditions: communication only task (com task), multi-task low workload (multi-task low), multi-task medium workload (multi-

task med), and multi-task high workload (multi-task high). The trial length for each condition was 3 min.

The com task condition required monitoring of the communication task only. All other tasks (light detection, gauge monitoring, and tracking) were not monitored and maintained a normal state (no fault events or tracking deviations). There were a total of 8 target communication events and 2 noise events. Each event was separated by at least 7 secs.

The multi-task conditions required monitoring of all tasks (communication, light detection, gauge monitoring, and tracking). All of the multi-task conditions contained a total of 10 communication events (8 target, 2 noise). The multi-task conditions varied in the number of fault events (light detection and gauge monitoring) within a trial, the time interval between fault events and communication events, and tracking difficulty. The multi-task low condition contained 4 light detection events (2 green offset and 2 red onset), 4 gauge faults, and a low difficulty disturbance function for tracking. Each event was separated by at least 9 secs. The multi-task med condition contained 12 light detection events (6 green offset and 6 red onset), 10 gauge faults, and a medium difficulty disturbance function for tracking. Each event was separated by at least 3 secs. The multi-task high condition contained 16 light detection events (8 green offset and 8 red onset), 18 gauge faults, and a high difficulty disturbance function for tracking. Each event was separated by at least 2 secs.

Subjects were instructed to treat tracking as the primary task and the communication as the secondary task. They were told to time share these tasks with the light detection and gauge monitoring tasks. The communication, light detection, and gauge monitoring tasks were to be performed as quickly and accurately as possible. Accuracy (false alarms and misses) RTs and tracking root mean square (RMS) were recorded. A miss was recorded if subjects did not respond to the communication task, light detection, or gauge fault within 15 sec. A false alarm was recorded if the incorrect communication command, light, or gauge was selected.

Subjects completed six sessions. Session 1 and 6 required both electrophysiological (EEG, EKG, and respiration), and behavioral data collection. In these sessions, subjects completed a total of 24 trials (6 trials of the com-task, multi-task low, multi-task med, and multi-task high). In addition, subjects completed the subjective workload rating scales after the last 3 trials which consisted of the multi-task low, med, and high difficulty conditions (randomly ordered). Sessions 2 through 5 were training sessions during which only behavioral data was recorded. In these sessions, subjects completed a total of 30 trials (10 trials of multi-task low, multi-task med, and multi-task high). The order of task conditions in each session was random with the constraint that no two of the same conditions occurred in sequence. Session 1 and 6 took approximately 3 hours (60 min to hook up electrophysiology apparatus and 90-120 min to perform task). Sessions 2 through 5 took



approximately 90-100 minutes. Subjects took a 1-2 min break after each trial and a 10-15 min break midway through each session.

### Results and Conclusions

Results and conclusions are pending data collection completion and analysis.

## References

- Kaufman, L., Curtis, S., Wang, J. Z., Williamson, S. J. (1991). Changes in cortical activity when subjects scan memory for tones. Electroencephalography & Clinical Neurophysiology, 82, 266-284.
- van Winsum, W., Sergeant, J. A., & Geuze, R. H. (1984). The functional significance of event related desynchronization of alpha rhythm in attentional and activating tasks. Electroencephalography & Clinical Neurophysiology, 58, 519-524.
- Yeh, Y. Y., and Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. Human Factors, 30, 111-120.
- Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalography & Clinical Neurophysiology, 42, 817-826.
- Pfurtscheller, G., & Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. Electroencephalography & Clinical Neurophysiology, 46, 138-146.

- Pfurtscheller, G., & Klimesch, W. (1989). Cortical activation pattern during reading and semantic classifications studied with dynamic ERD mapping. In: Topographic Brain Mapping of EEG and Evoked Potentials, Maurer K., ed. Berlin: Springer, pp. 303- 313.
- Geuze, R., & van Winsum, W. (1987). Event-related desynchronization and P300. Psychophysiology, 24, 272-277.
- J. R. Comstock and R. J. Arnegard (1992). The Multi-attribute task battery for human operator workload and strategic behavior research. NASA Technical Memorandum 104174.
- Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography & Clinical Neurophysiology, 10, 371-375.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

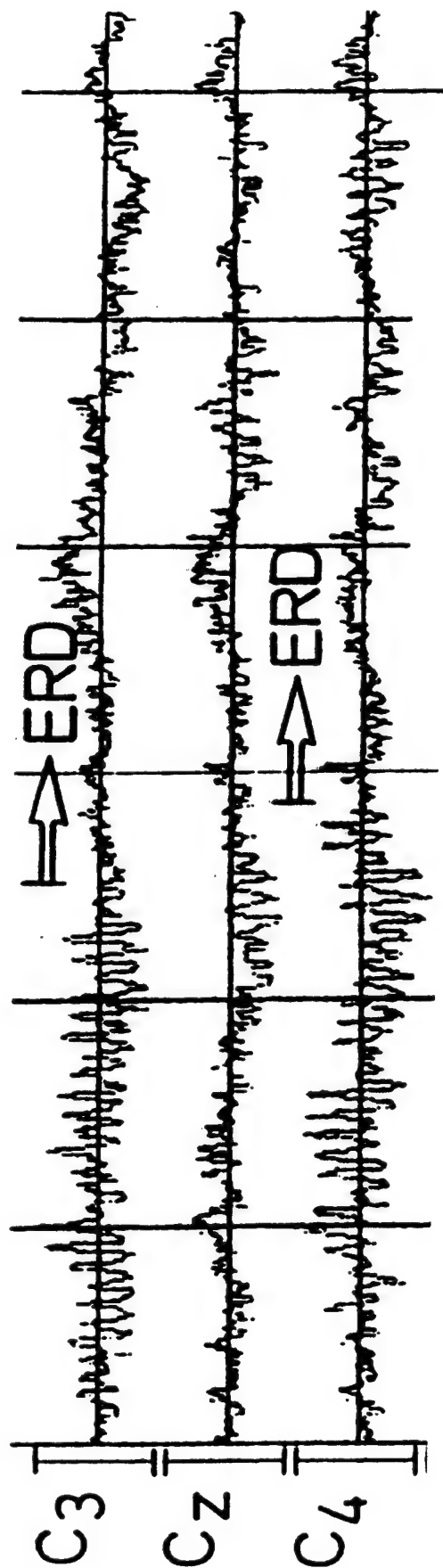


Figure 1. An example of alpha ERD over central scalp sites.

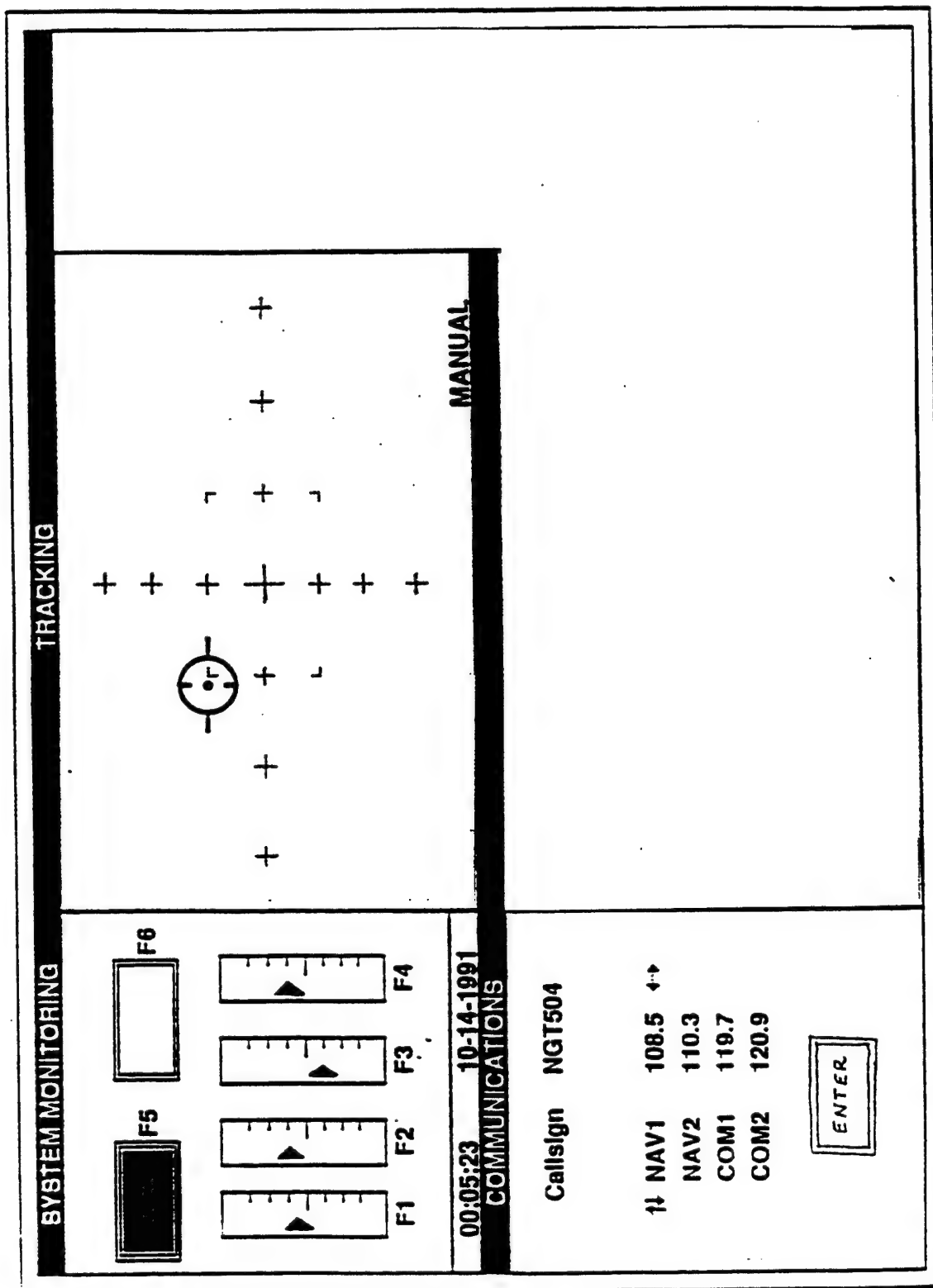
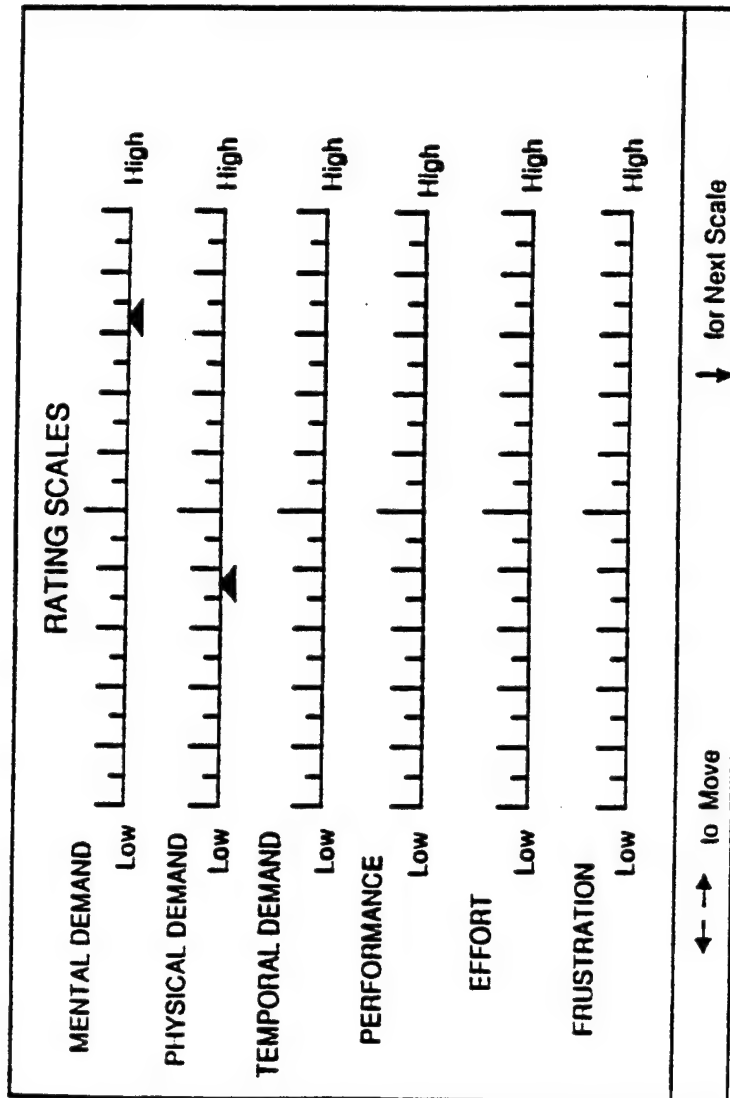


Figure 2. MATB visual display



**Figure 3. NASA TLX Rating Scale**

**DEVELOPMENT OF A GAS UPTAKE TECHNIQUE TO DETERMINE KINETIC CONSTANTS FOR  
METABOLISM OF INHALED TRICHLOROETHYLENE BY MALE B6C3F1 MICE**

**Marc S. Greenberg  
Graduate Student  
Biomedical Sciences Ph.D. Program  
Department of Pharmacology & Toxicology**

**Wright State University  
Dayton Ohio, 45435**

**Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory**

**Air Force Office of Scientific Research  
Bolling Air Force Base  
Washington, DC**

**and**

**Armstrong Laboratory  
Toxicology Division  
Wright-Patterson Air Force Base  
Dayton, Ohio**

**September, 1995**

DEVELOPMENT OF A GAS UPTAKE TECHNIQUE TO DETERMINE KINETIC CONSTANTS FOR  
METABOLISM OF INHALED TRICHLOROETHYLENE BY MALE B6C3F1 MICE

Marc S. Greenberg  
Graduate Student  
Biomedical Sciences Ph.D. Program  
Department of Pharmacology & Toxicology  
Wright State University

Abstract

Methods were developed to investigate the uptake of the chlorinated hydrocarbon trichloroethylene (TCE) in small numbers of B6C3F1 mice. A gas uptake system was constructed in order to measure the removal of the compound from a recirculating atmosphere by mice exposed to 100, 300, 600, 1000 and 5000 ppm TCE in air. These data will be used in the development of a physiologically-based pharmacokinetic (PB-PK) model to describe the uptake of TCE and its metabolism to active metabolites in mice. Metabolic rate constants needed for the PB-PK will be estimated by a computer simulation approach using experimental gas uptake results as input data. Following a calibration period, it was determined that the gas uptake system is ready for use since it contains no leaks, a linear standard curve over a broad range of TCE concentrations (100-7000 ppm) has been constructed and the loss rate from an empty chamber remains constant over this range of concentrations. Simulations of planned exposure concentrations and preliminary experiments with mice exposed to 1000 ppm TCE indicate that the maximum rate of metabolism,  $V_{\max}$ , is between 4.0 and 32.7  $\text{mg}\cdot\text{kg}^{-1}\cdot\text{h}^{-1}$  for male B6C3F1 mice. Completion of the study at all TCE concentrations described in the experimental design will give an accurate estimate of this metabolic constant. The implications of obtaining an accurate value for  $V_{\max}$  over a broad range of TCE concentrations to the PB-PK modeling effort are addressed in the discussion.



# DEVELOPMENT OF A GAS UPTAKE TECHNIQUE TO DETERMINE KINETIC CONSTANTS FOR METABOLISM OF INHALED TRICHLOROETHYLENE BY MALE B6C3F1 MICE

Marc S. Greenberg

## Introduction

Trichloroethylene (TCE) a widely produced chlorinated solvent, degreasing agent and extractant is a persistent environmental contaminant that occurs in workplaces, air, groundwater, soil and foods (Malatoni et al., 1988). Short term exposure to TCE results in mild central nervous system effects such as dizziness, sleepiness, facial numbness and irritation to the eyes, nose and throat (ATSDR, 1989). TCE has been shown to induce cancer in rodents during laboratory bioassays and thus has been classified as a B2 carcinogen by the U.S. Environmental Protection Agency (USEPA, 1985; NTP, 1990). Human exposure to TCE in the environment primarily occurs from contaminated groundwater or from occupational exposure (Fisher and Allen, 1993). Equivocal evidence that TCE is a human carcinogen was demonstrated in separate cohort mortality studies of workers from a U.S. aircraft maintenance facility and a Swedish manufacturing plant (Spirtas et al., 1991; ATSDR, 1993; Axelson et al., 1994). Evidence suggests that metabolic activation of parent trichloroethylene is required for toxicity and carcinogenicity in the liver and in extrahepatic organs such as the lung, the liver and the kidney (Bergman, 1983; Walles, 1986; Larson and Bull, 1992). Since dose-response relationships have been drawn between total TCE metabolism and toxicity, a physiologically-based pharmacokinetic (PB-PK) model based upon a collected database can be useful for predicting the toxicity and carcinogenicity of TCE to mammals, including humans (Dallas et al., 1991; Allen and Fisher, 1993).

A PB-PK with a large database containing necessary physiological parameters (e.g., the metabolic constants  $V_{max}$  and  $K_m$ , cardiac output, blood flows to specific organs, respiration, etc.) exists for B6C3F1 mice (Fisher et al., 1991) exposed to TCE and related contaminants in gas uptake systems and through other routes of administration. With proper scaling factors based on allometric relationships, dose extrapolation, and hence, the carcinogenic potential of a compound can be predicted from an existing B6C3F1 mouse-derived PB-PK to other mammals (e.g., rats, humans, etc.) (Koizumi, 1989). However, while the Fisher et al. (1991) model adequately describes the kinetics of TCE metabolism in female B6C3F1 mice, the data for males is highly variable and requires further investigation. With a more accurate description of the metabolic constants in males of this strain, a more comprehensive, biologically-based description of TCE toxicity in mammals could then reduce or eliminate the need for costly cancer bioassays involving large numbers of vertebrate animals.

## Methods

Gas uptake studies using a recirculating atmosphere in a closed exposure chamber (Filsler and Bolt, 1981) allow the investigator to estimate Michaelis-Menten type metabolic kinetic constants ( $V_{max}$  and  $K_m$ ) for the *in vivo* metabolism of inhaled TCE. These physiological constants are necessary in a PB-PK model. Because the Fisher et al. (1991) study based its estimates of kinetic constants on TCE exposures of 14 male or

female mice, the investigator of the current study reasoned that replicate experiments using a smaller number of animals in an exposure chamber would give less variable results and thus better model predictions of TCE metabolism in male B6C3F1 mice.

*Gas uptake system.* The closed atmosphere exposure system is a modified version of the system described in Gargas et al. (1986b). It consists of a 2.8 L desiccator jar with gas inlet and outlet fittings built into the lid. A small amount (25 g) of barilyme is placed in the bottom of the jar to remove the CO<sub>2</sub> produced by respiration of the mice. Mice are placed into the jar on a desiccator tray which lies 3 cm above the layer of barilyme and the chamber is sealed with a thin layer of vacuum grease (Dow Chemical Co., Midland, MI). The atmosphere is recirculated with a stainless-steel metal bellows pump (2.0 L/min) (Metal Bellows Corporation, Sharon, MA) and the water vapor produced by respiration of the mice is removed by an inline moisture trap consisting of a sealed glass cylinder (250 mL) containing 30 glass marbles (1 cm diam. each), placed into an ice bath. Oxygen is measured with an oxygen monitor (MDA Scientific, Inc., Lincolnshire, IL) and is added to the system as needed to maintain a concentration of 21%. The atmospheric pressure in the chamber is continuously monitored with a magnahelic gauge (F. W. Dwyer Mfg. Co., Michigan City, IN) and is maintained at -2 to +2 inches of water. All flow lines to the system are comprised of stainless steel and TCE (99+% pure, spectrophotometric grade, Aldrich Chemical Co., Midland, MI) is introduced as a liquid through a rubber septum fitted 10 inches upstream of the chamber. A heat gun is used to rapidly vaporize the injected chemical in order to achieve a gaseous initial concentration. The disappearance of TCE from the exposure atmosphere is measured by gas-liquid chromatography, using an automated gas chromatograph system described below.

The concentration of TCE in the exposure chamber is analyzed every 5 minutes after injection of the chemical for the first 30 min and then every 15 min for the duration of exposure (6 h). An automatic gas sampling valve (1.0 mL), programmed by a sampler/event control module (HP 19405A, Hewlett-Packard, Avondale, PA) and an integrator (HP 3396A), is used to introduce chemical onto the chromatographic column. An inlet and an outlet line connect the programmable sampling valve to the exposure chamber, and this loop is set at a flow rate of 100 mL/min. The chromatograph (HP 5890 Series II) is equipped with a hydrogen flame ionization detector (FID) with helium used as the carrier gas (30 mL/min). Analyte detection is performed on a 12' x 1/8" i.d. stainless steel column packed with 10% SE-30 on 80/100 mesh Chromosorb W-HP (Supelco, Inc., CA). The detector and sample valve area temperatures are maintained at 250° C and 175° C, respectively. The column oven temperature is held constant at 140° C.

*Animal exposure to TCE.* Male and female B6C3F1 mice (age, 4-6 weeks), weighing 25-30 g and virus antibody-free, will be obtained from Charles River Breeding Laboratories (Kingston, NY). Animals will be allowed free access to Purina chow food and water and will be housed over autoclaved bedding, in American Association for Accreditation of Laboratory Animal Care-accredited facilities at Wright-Patterson Air Force Base, Dayton, for at least 5 days before use.

Three male B6C3F1 mice are used for each exposure and exposure duration is 6 h. Five initial concentrations of TCE will be tested: 100, 300, 600, 1000, and 5000 ppm. These exposure levels were chosen in order to observe kinetic behavior over a wide concentration range. Unexposed animals will be kept in the gas uptake chamber for an equivalent amount of time. Each exposure concentration and the control mice will be tested a minimum of 3 times. A nonspecific, first-order loss rate from an empty chamber, which is independent of concentration, was measured for each level of TCE and will be accounted for in the data analysis. The maximum rate of metabolism ( $V_{\max} \pm \text{SD}$ ;  $\text{mg}\cdot\text{kg}^{-1}\cdot\text{h}^{-1}$ ) and the Michaelis constant ( $K_m$ ;  $\text{mg}\cdot\text{L}^{-1}$  blood) for a mouse (allometrically scaled) will be estimated by computerized nonlinear least-squares techniques (Simusolv, Dow Chemical Co.) using a five-compartment simulation model (Ramsey and Andersen, 1984; Gargas et al., 1986b). Four physiologically-based compartments represent the mouse (e.g., liver, blood, slowly perfused tissue such as muscle, and rapidly perfused tissue such as kidney and brain) and one compartment represents the gas uptake chamber. The experimentally-derived values for the kinetic constants can then be used in predictive models of TCE toxicity for describing exposures by inhalation and other routes of administration. In addition, the animals used in gas uptake exposures will be dissected and their tissues analyzed for TCE and its metabolites using the methods described below. Tissue analysis will be carried out in order to incorporate metabolite data into the current PB-PK model for mice (Fisher et al., 1991) at some future date.

*Tissue sampling and metabolite analysis.* Mice will be immediately sacrificed by overexposure to  $\text{CO}_2$  following a 6 h gas uptake exposure and their tissues dissected for analysis of parent TCE and its metabolites. From each mouse the following tissues will be removed, placed into pre-weighed, 4-mL screw-cap vials with teflon coated septa containing the proper extraction solvents (see below), homogenized and then flash frozen with liquid nitrogen until analysis: 1) 4 x 0.2 g of liver, kidney, and fat, 2) 4 x 0.2 mL of blood, and, 3) the lungs of all 3 mice will be pooled and then divided into 4 equal portions of tissue. From 1), 2) and 3) above, each portion of tissue collected will be analyzed separately for the following compound(s): a) TCE and chloral (CL), b) free chloral hydrate (CH) and trichloroethanol (TCOH), c) total TCOH (i.e. to distinguish TCOH from TCOH-Glucuronide), and, d) di- and tri- chloroacetic acids (DCA) and (TCA). All reagents used for extractions of TCE and its metabolites from tissues in the experiments will be supplied by Fisher Scientific Co., and will be of HPLC-grade quality or better. Extraction procedures and gas chromatographic (GC) analyses have been adapted from many sources (Odum et al., 1992; Templin et al., 1993; *unpublished laboratory SOP*) and thus will only be described in brief.

TCE and CL will be extracted into *tert*-butyl methyl ether and analyzed by GC with an electron capture detector (ECD) by liquid injection. Free CH and free TCOH will be extracted into ethyl acetate and analyzed by GC-ECD by hand injection. Conjugated TCOH (TCOH-G) will be released with sulfuric acid followed by extraction of total TCOH into hexane prior to injection for analysis by GC-ECD. TCA and DCA will be extracted into ether, reacted with diazoethane and analyzed by GC-ECD following liquid injection. For

these analyses an HP 5890 Series II GC equipped with a 30 m x 0.53 Vocol capillary column (Supelco, Inc., CA) will be used. The carrier and make up gasses will be 95% Argon/5% Methane and will be set to 6 mL/min and 25 mL/min, respectively. The injection port and ECD detector temperatures will be 175° C and 300° C, respectively. For TCE and CL analysis the column oven temperature will be set to 50° C and run for 15 min. For detection of free CH, free and total TCOH, TCE and DCA the initial oven temperature will be set to 80° C for 4 min and ramped at 25 ° C/min to 170 ° C and held for 2.2 min. The total run time will be 10 min.

Concentrations of metabolites will be reported as mean ( $\pm 1$  SE) in units of nmol/g tissue. Analysis of variance (ANOVA; SAS, 1989) will be used to compare the response levels of each TCE analyte in each tissue type sampled from the mice exposed to contaminated air. Differences in tissue concentration values due to the effect of TCE concentration will be considered significant if  $p \leq \alpha = 0.05$ .

## Results

At the time of the preparation of this report, the following portions of the ongoing research objectives outlined above were completed: 1) set up and calibration of the gas uptake system, 2) initial model simulations using recently measured tissue partition coefficients for TCE and B6C3F1 mice (*R. Abbas, unpublished data*), and 3) two trial gas uptake runs with live mice at an exposure concentration of 1000 ppm TCE in air.

The gas uptake system has been calibrated to a TCE concentration range of 100-7000 ppm in air. All lines have been checked for leaks and the loss of TCE from an empty chamber over 6 h was measured at each concentration to insure that the first order loss rate was less than  $0.05 \text{ h}^{-1}$ . The mean ( $\pm 1$  SD) loss rate for the concentrations used to construct a linear standard curve for TCE (Fig. 1) was  $0.028 \pm 0.006 \text{ h}^{-1}$ . Two loss runs were performed at 1000 ppm TCE

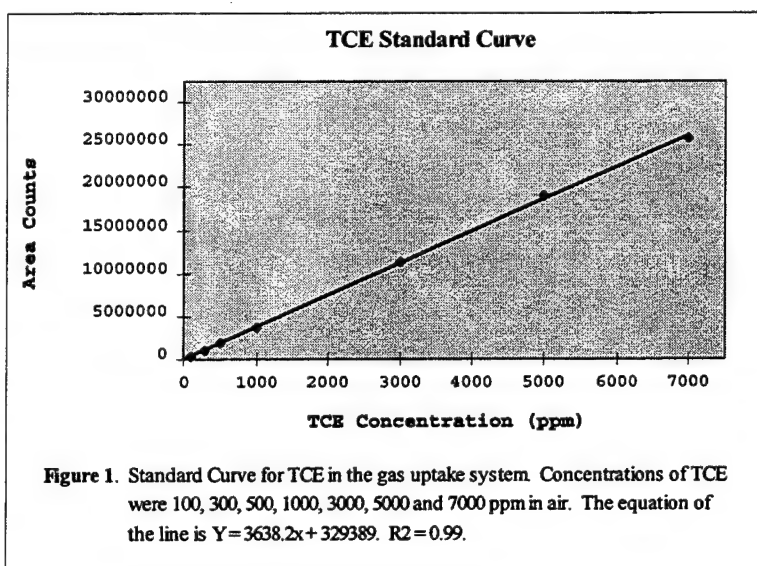
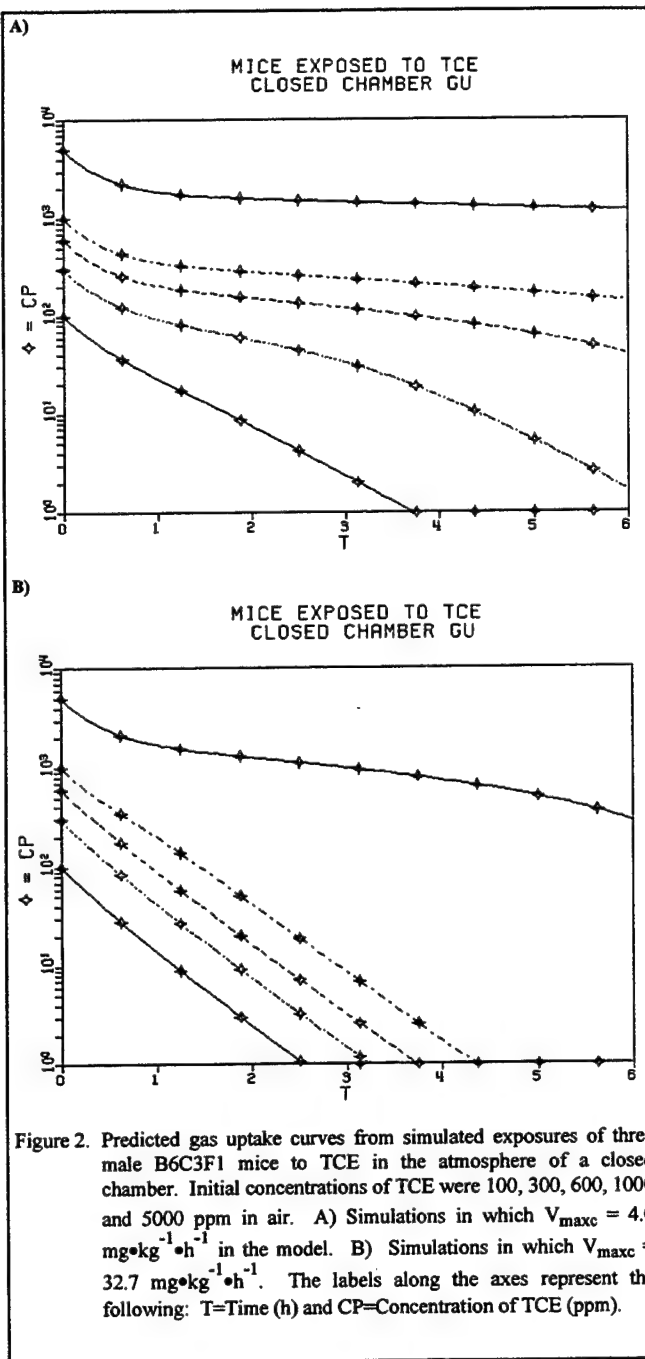


Figure 1. Standard Curve for TCE in the gas uptake system. Concentrations of TCE were 100, 300, 500, 1000, 3000, 5000 and 7000 ppm in air. The equation of the line is  $Y = 3638.2x + 329389$ .  $R^2 = 0.99$ .

using three dead B6C3F1 mice in order to investigate the possibility that some TCE loss during an experiment may be due to the partitioning of the compound to the external body parts or the fur of the test subjects. The loss rate in both dead-mice runs was  $0.025 \text{ h}^{-1}$ , which was within the range of the loss rates described above.

Initial simulations were run at the 5 exposure concentrations and only the parameter of  $V_{\text{maxc}}$  was changed between the simulation results shown in Figure 2. When  $V_{\text{maxc}}$  was set to  $4.0 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ , the model predicted that, within 6 h, TCE in the atmosphere would be completely depleted by three mice at an initial exposure concentration of 100 ppm (Fig. 2a). Fisher et al. (1991) reported a simulation-optimized  $V_{\text{maxc}}$  of 32.7



$\text{mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$  for male B6C3F1 mice exposed to TCE in a 9 L gas uptake system. When this value was applied to the current gas uptake system, the model predicted that TCE would be completely removed from the atmosphere by three mice at all initial exposure concentrations except for 5000 ppm TCE (Fig. 2b). Figure 3 shows the results of an exposure of three mice to 1000 ppm TCE in the gas uptake chamber compared to simulated results when the model was run with  $V_{maxc} = 4.0 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ . The model slightly underpredicts the removal of TCE from the recirculating atmosphere by the mice.

### Discussion

The gas uptake system is ready to use for the replicate experiments outlined in the preceding methods section. A linear, reproducible standard curve for TCE from 100-7000 ppm (Fig. 1) insures that accurate concentration measurements of the compound at each time point will be made by the investigator. Results from the system calibration period indicate that any loss of TCE from the atmosphere above 2.5% per hour will be due to the uptake and metabolism of TCE by the mice and not due to the compound partitioning to the glass of the chamber, the tubing, or the fur of the mice. This assumption is important to the PB-PK

modeling effort that will be based on data generated by the exposure of mice via inhalation of gaseous TCE in a closed chamber.

Simulations of the uptake of TCE by mice appear to be very sensitive to the maximum rate of metabolism,  $V_{maxc}$  (Fig. 2). Data from a single gas uptake exposure (Fig. 3) indicate that a  $V_{maxc}$  that is much less than  $32.7 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$  (Fisher et al., 1991) but slightly greater than  $4.0 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$  will most accurately estimate the uptake curves of TCE-exposed B6C3F1 mice. This may be explained by the difference in methodology between the Fisher et al. (1991) study and the current investigation. In the former, a larger

exposure chamber (9 L) and a greater number of mice (15 males) were used for an exposure. This may have led to more variable data than can be obtained by using a smaller chamber with fewer subjects. In gas uptake studies with B6C3F1 mice exposed to chloroform, only one subject was used in an exposure (Gearhart et al., 1993). Gargas et al. (1986a & b) used three rats in gas uptake studies of over 30 halogenated compounds. Perhaps, the use of three test subjects, as in the current experimental design will resolve this potential discrepancy between  $V_{maxc}$  estimates for male B6C3F1 mice. However, the observation that all the TCE is taken up by mice except at the highest concentration when simulations were run with

$V_{maxc} = 32.7 \text{ mg}\cdot\text{kg}^{-1}\cdot\text{h}^{-1}$  (Fig. 2b), should concern the investigator in that if this rate constant is real, then only at very high TCE concentrations will  $V_{maxc}$  and  $K_m$  be reliably estimated. The outcome of the trial experiment (Fig. 3) suggests that this value is too high. Future research results, especially from lower concentration exposures, will clarify this issue.

The complete results of the ongoing investigation should give a more accurate assessment of the uptake and metabolism of TCE in male B6C3F1 mice. This should lead to better optimized values of  $V_{maxc}$  from the PB-PK computer model. The importance of obtaining a more accurate estimate of this metabolic constant is twofold. First, this information will lead to better interpretation of the tissue metabolite data to be collected following the exposures. Second, a successful PB-PK model for TCE in mice, that works over a broad range of exposure concentrations, would be more easily developed. This will, ultimately, lead to better dose extrapolations and hence more accurate estimates of the health risks to humans that are exposed to TCE.

#### Literature Cited

- Allen, B. C. and J. W. Fisher. 1993. Pharmacokinetic modeling of trichloroethylene and trichloroacetic acid in humans. *Risk Anal.* 13(1): 71-86.
- ATSDR. 1989. Toxicological profile for trichloroethylene. U. S. Public Health Service and USEPA. ATSDR/TP-88/24. 139 pp.
- ATSDR. 1993. Toxicological profile for trichloroethylene (update). U. S. Public Health Service and USEPA. ATSDR/TP-92/19. 126 pp.

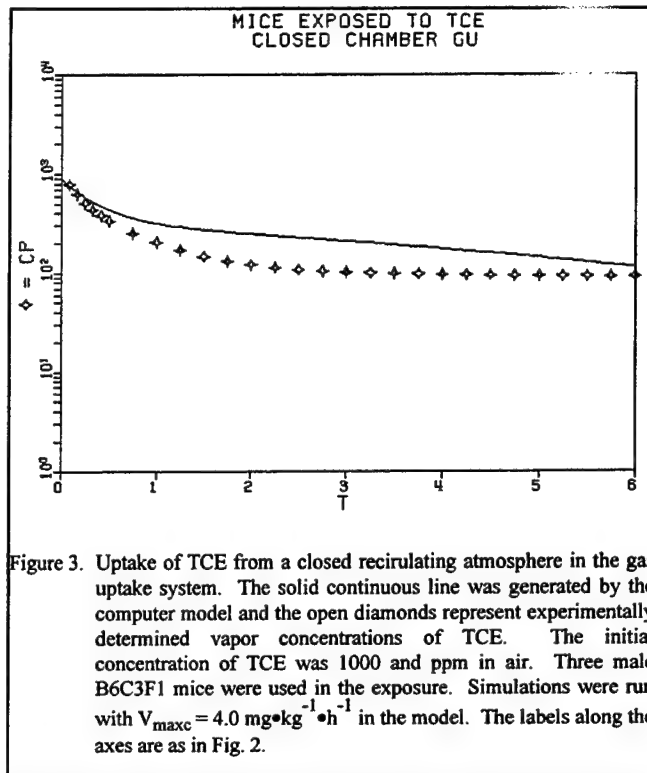


Figure 3. Uptake of TCE from a closed recirculating atmosphere in the gas uptake system. The solid continuous line was generated by the computer model and the open diamonds represent experimentally determined vapor concentrations of TCE. The initial concentration of TCE was 1000 and ppm in air. Three male B6C3F1 mice were used in the exposure. Simulations were run with  $V_{maxc} = 4.0 \text{ mg}\cdot\text{kg}^{-1}\cdot\text{h}^{-1}$  in the model. The labels along the axes are as in Fig. 2.

- Axelsson, O., A. Selden, K. Andersson and C. Hogstedt. 1994. Updated and expanded Swedish cohort study on trichloroethylene and cancer risk. *J. Occup. Med.* 36(5): 556-562.
- Bergman, K. 1983. Interactions of trichloroethylene with DNA in vitro and with RNA and DNA of various mouse tissues in vivo. *Archiv. Toxicol.* 54: 181-193.
- Dallas, C. E., J. M. Gallo, R. Ramanathan, S. Muralidhara and J. V. Bruckner. 1991. Physiological pharmacokinetic modeling of inhaled trichloroethylene in rats. *Toxicol. Appl. Pharmacol.* 110: 303-314.
- Filser, J. G. and H. M. Bolt. 1981. Inhalation pharmacokinetics based on gas uptake studies. I. Improvement of kinetic models. *Arch. Toxicol.* 47: 279-292.
- Fisher, J. W. and B. C. Allen. 1993. Evaluating the risk of liver cancer in humans exposed to trichloroethylene using physiological models. *Risk Anal.* 13(1): 87-95.
- Fisher, J. W., M. L. Gargas, B. C. Allen and M. E. Andersen. 1991. Physiologically based pharmacokinetic modeling with trichloroethylene and its metabolite, trichloroacetic acid in the rat and mouse. *Toxicol. Appl. Pharmacol.* 109: 183-195.
- Gargas, M. L., Clewell, H. J., III, and Andersen, M. E. 1986a. Metabolism of inhaled dihalomethanes *in vivo*: Differences of kinetic constants for two independent pathways. *Toxicol. Appl. Pharmacol.* 82: 211-223.
- Gargas, M. L., M. E. Andersen and H. J. Clewell. 1986b. A physiologically based simulation approach for determining metabolic constants from gas uptake data. *Toxicol. Appl. Pharmacol.* 86: 341-352.
- Gearhart, J. M., Seckel, C. and Vinegar, A. 1993. *In vivo* metabolism of chloroform in B6C3F1 mice determined by the method of gas uptake: The effects of body temperature on tissue partition coefficients and metabolism. *Toxicol. Appl. Pharmacol.* 199: 258-266.
- Koizumi, A. 1989. Potential of physiologically based pharmacokinetics to amalgamate kinetic data of trichloroethylene and tetrachloroethylene obtained in rats and man. *Brit. J. Indust. Med.* 46: 239-249.
- Larson, J. L. and R. J. Bull. 1992. Metabolism and lipoperoxidative activity of trichloroacetate and dichloroacetate in rats and mice. *Toxicol. Appl. Pharmacol.* 115: 268-277.
- Maltoni, C., G. Lefemine, G. Cotti and G. Perino. 1988. Long-term carcinogenicity bioassays on trichloroethylene administered by inhalation to Sprague-Dawley rats and Swiss and B6C3F1 mice. *Ann. NY Acad. Sci.* 354: 316-342.
- NTP. 1990. Toxicology and carcinogenesis studies of trichloroethylene (without epichlorohydrin) (CAS No. 79-01-6) in F344/N rats and B6C3F1 mice (gavage studies). U. S. Department of Health and Human Services. Technical Report Series No. 243. NIH Publ. No. 90-1779, Research Triangle Park, NC.
- Odum, J., J. R. Foster and T. Green. 1992. A mechanism for the development of Clara cell lesions in the mouse lung after exposure to trichloroethylene. *Chem.-Biol. Interact.* 83: 135-153.
- Ramsey, J. C. and M. E. Andersen. 1984. A physiologically based description of the inhalation pharmacokinetics of styrene in rats and humans. *Toxicol. Appl. Pharmacol.* 73: 159-175.

- Spiro, R., P. A. Stewart, J. S. Lee, D. E. Marano, C. D. Forbes, D. J. Grauman, H. M. Pettigrew, A. Blair, R. N. Hoover and J. L. Cohen. 1991. Retrospective cohort mortality study of workers at an aircraft maintenance facility. I. Epidemiological results. *Brit. J. Indust. Med.* 48: 515-530.
- Templin, M. V., J. C. Parker and R. J. Bull. 1993. Relative formation of dichloroacetate and trichloroacetate from trichloroethylene in male B6C3F1 mice. *Toxicol. Appl. Pharmacol.* 123: 1-8.
- USEPA. 1985. Health assessment document for trichloroethylene. Washington, D. C., Office of Health and Environmental Assessment.
- Walles, S. A. S. 1986. Induction of single-strand breaks in DNA of mice by trichloroethylene and tetrachloroethylene. *Toxicol. Lett.* 31: 31-35.



THE EFFECTS OF INDIVIDUALIZED GOALS  
ON COMPLEX TASK PERFORMANCE

Jennifer L. Greenis  
Graduate Student  
Department of Psychology

Michigan State University  
129 Psychology Research Building  
East Lansing, MI 48824-1117

Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1995

# THE EFFECTS OF INDIVIDUALIZED GOALS ON COMPLEX TASK PERFORMANCE

Jennifer L. Greenis  
Graduate Student  
Department of Industrial and Organizational Psychology  
Michigan State University

## Abstract

Despite the beneficial effects of goal setting on simple tasks, research has shown that goal setting on complex or novel tasks may be ineffective or even detrimental to performance (Earley, Connolly, & Ekegren, 1989; Kanfer & Ackerman, 1989). In order to balance the beneficial effects of goals on motivation with the potentially harmful effects on self-efficacy, affect, attention, and strategy, it is recommended that goals be designed such that individuals evidence moderate goal-performance discrepancies (Bandura & Cervone, 1983, 1986), and progression toward their goals (Hsee & Ableson, 1991). Based on these two guidelines, the present study investigated the effects of individualizing goals by tailoring goals to an individual's performance rather than the group's performance, and by updating goals over time. A three x three (goal type x block) repeated measures factorial design was used. Subjects were assigned to one of three goal conditions: do-your-best (control group), standard (group-based, non-updated), or individualized (individual-based, updated) goals. Performance was assessed across three blocks of an air traffic control task, TRACON. Each block consisted of five trials. Prior to performance assessment, individuals were given two practice trials in which they were told to "do your best." Goals were assigned during the three performance blocks. Individuals in the control group were told to "do your best" throughout the entire experiment. Individuals in the standard goal condition were assigned one goal, equivalent to the 85th percentile of the control group's performance. Subjects in the individualized goal condition were given goals which were tailored to their individual capabilities and performance levels, and updated over time. Individualized goals were calculated by assessing each individual's initial performance, predicting his or her upcoming performance using regression (based on control group data), and then incrementing his or her predicted performance by one standard deviation. Following each block of trials, various cognitions (self-efficacy, negative affect, etc.) were assessed. Fifty-six subjects participated in the experiment. Results indicated that individuals given "do-your-best" goals consistently outperformed subjects given standard goals; however, this effect was only statistically significant during the first block of performance. Although subjects given individualized goals outperformed those given standard goals during the first block, they performed worse than individuals assigned "do-your-best" or standard goals during the final two blocks. Implications of these findings are discussed.

# THE EFFECTS OF INDIVIDUALIZED GOALS ON COMPLEX TASK PERFORMANCE

Jennifer L. Greenis

## Introduction

The impact of rapid technological developments and diverse consumer demands on today's workplace has created many dynamic jobs which require frequent training to update skills (Goldstein & Gilliam, 1990; Howell & Cooke, 1989). Given the novel and complex nature of these jobs and the tasks involved, sustained task attention and extensive practice are necessary for successful skill acquisition. However, in order for learners to sustain task attention when learning a complex task (especially when facing task difficulties and failure), they must be motivated. A learner's level of motivation is influenced by many factors during skill acquisition. Two factors which have received a great deal of attention are goals and feedback (Bandura, 1986; Dweck, 1986).

Goal setting has proven to be an extremely effective motivational intervention to aid skill acquisition and performance on simple tasks such as performing number comparison tasks or solving arithmetic problems (Erez, 1977; Matsui, Okada, & Inoshita, 1983). Research on simple tasks has consistently demonstrated that (1) when individuals are committed to a goal and have the ability to achieve it, more difficult goals lead to better performance; and (2) specific, challenging goals lead to better performance than vague, easy, or no goals (Locke, Shaw, Saari, & Latham, 1981; Mento, Steel, & Karren, 1987). Despite the beneficial effects of goal setting on simple tasks, research has shown that goal setting on complex or novel tasks may be ineffective or even detrimental to performance (Earley, Connolly, & Ekegren, 1989; Kanfer & Ackerman, 1989). When researchers have manipulated task complexity on goal setting tasks, they have found that specific, difficult goal assignments positively affect performance in low complexity versions of a task, but not in complex task versions (Johnson & Kanfer, 1992; Wood, Bandura, & Bailey, 1990). In line with these studies, a meta-analysis by Wood, Mento, and Locke (1987) revealed that the magnitude of goal effects on performance decreases as the complexity of the task increases.

## Discussion of the Problem

In order to design performance goals which will aid complex task performance, the mechanisms by which goals impact motivation and performance will first be reviewed. Following, several theories posited to explain why goals are less effective on complex tasks will be discussed, and guidelines for goal assignment on complex tasks will be provided.

Motivational and Cognitive Mechanisms of Goal Setting. Activation of self-regulatory processes is expected when individuals adopt difficult goals and believe that they have the ability to accomplish those goals. Bandura (1982) and Kanfer and Ackerman (1989) assert three main components of self-regulation: self-monitoring, self-evaluation, and self-reactions. Self-monitoring refers to attention given to specific aspects of one's own behavior. Self-monitoring typically leads to self-evaluation (i.e., the comparison between one's standard and relevant performance). In order for these self-regulatory processes to be enacted, feedback which provides learners with

information about the consequences of their actions is necessary. Attention to feedback induces self-evaluation by providing a referent in the goal-performance comparison (and identifying goal-performance discrepancies). The evaluative processes in turn influence an individual's affective reactions to the task. These self-reactions include satisfaction and self-efficacy judgments.

Based on this model of self-regulation, three direct motivational mechanisms (direction of attention, effort, and persistence) and one indirect cognitive mechanism (strategy development) are thought to be responsible for the effects of goal setting on performance (Locke et al., 1981). First, goals direct activity toward goal-relevant activities. Second, they regulate effort by prompting individuals to adjust their effort according to the goal difficulty. Third, goals increase persistence by inducing people to work longer. All three of these mechanisms are direct, automatic consequences of setting goals; however, at times, they may be insufficient for goal attainment. In such situations (e.g., on complex tasks), individuals must develop or find new strategies which will lead to goal attainment. Strategy development is considered an indirect goal mechanism because it is not an automatic, well-learned process.

The Effects of Goal Setting on Complex Task Performance. Overall, goal setting on simple tasks improves performance by encouraging learners to engage in self-regulatory processes. However, it appears that goal setting may actually hurt performance on complex tasks as a result of increased self-regulation (Earley, Connolly, & Ekegren, 1989; Kanfer & Ackerman, 1989). Several explanations have been proposed to explain this contradictory finding. Researchers assert that goal setting on complex or novel tasks leads to poor performance because goals stimulate (1) competition for limited cognitive resources (Kanfer & Ackerman, 1989), or (2) ineffective strategy development and switching (Earley, Connolly, & Ekegren, 1989), or (3) negative affect, reduced goal commitment, and task withdrawal as a result of large negative goal-performance discrepancies (Carver & Scheier, 1981).

The Resource Allocation model asserts that goal setting activates self-regulatory processes which utilize valuable cognitive resources (Kanfer & Ackerman, 1989). When the attentional demands of a task are high (i.e., for complex or novel tasks), self-regulatory processes involved in performance evaluations and reactions may take away valuable resources necessary for task performance, and thereby hurt performance. The Strategy Model asserts that goal setting on complex tasks influences individual perceptions of performance pressure and self-efficacy judgments which directs strategy development and selection, and thereby affects performance (Earley, Connolly, & Ekegren, 1989). Lastly, Control Theory posits that the presence of negative goal-performance discrepancies (which are more likely to occur on complex or novel tasks) will create feelings of negative affect which may lead individuals to adjust their standards, effort, or attention to the task depending on their outcome expectancies (Carver & Scheier, 1981). Outcome expectancy assessments may be based on information pertaining to the situation or to an individual's internal qualities (i.e., self-efficacy). Theory and research need to address how performance goals can be designed (or assigned) to facilitate performance of complex or novel tasks. More specifically, how can task pressure, negative affect, and the usage of ineffective strategies be minimized, while self-efficacy is maximized? Additional goal setting research will now be reviewed and guidelines for the design of performance goals on complex tasks will be recommended.

Guidelines for Assigning Goals: Moderate Goal-Performance Discrepancies and Progression Toward the Goal. When individuals fail to meet their goals (i.e., when goal-performance discrepancies are large and negative), they typically become dissatisfied. The amount of dissatisfaction experienced can be influenced by the magnitude and direction of the perceived goal-performance discrepancy (Kanfer, 1990). Bandura and Cervone (1986) conducted an experiment to investigate the effects of goal-performance discrepancies (magnitude and direction) on self-reactions and performance. All eighty-eight subjects performed three sessions of a Schwinn AirDyne ergometer task. The first session served as practice, for which neither goals nor feedback were given. Prior to the second session, all subjects were given a goal to demonstrate at least a 50% increase in effort on this task. After the second session, subjects were either told that they (1) fell markedly short (of the goal), (2) fell moderately short, (3) fell minimally short, or (4) exceeded the standard. Subjects were randomly assigned to these four feedback conditions. Lastly, subjects performed a third session for which no goals or feedback were provided. Performance, self-satisfaction, self-efficacy for goal attainment, and self-set goals were assessed for each session. Results indicate that the largest increases in performance occurred in self-efficacious individuals who were given feedback indicating moderate goal-performance discrepancies. When small substandard goal-performance discrepancies occur, individuals tended to become complacent and decrease performance. *These findings suggest that goals should be designed such that they lead to moderate negative goal-performance discrepancies.*

Recent research indicates that dissatisfaction with performance is not only dependent on the magnitude and direction of one's goal-performance discrepancies, it is also dependent on one's progression toward the goal (Hsee & Ableson, 1991). That is, individuals who fall minimally short of a goal may still be dissatisfied when they do not demonstrate any progress toward their goal. Hsee and Ableson (1991) conducted two studies to investigate the effects of velocity (i.e., progression toward a goal) on satisfaction. In both studies, subjects rated their satisfaction with various goal-performance outcomes. Results indicate that individuals prefer progression toward a goal over no progression. In addition, individuals prefer no progression over performance degradation away from a goal. Furthermore, the more positive the velocity (the greater progression toward a goal), the greater one's satisfaction. Thus, when considering the effects of goals on self-regulation and performance, a dynamic view of outcomes and process approach to human motivation should be taken. People engage in behavior not just to seek a particular goal, but also to seek progression toward that goal over time. *These findings suggest that goals should be designed such that individuals will evidence progression toward their goals* (Hsee and Ableson, 1991).

Perceived negative goal-performance discrepancies between a standard and one's performance may lead individuals to question their own capabilities and experience feelings of negative affect. These dissatisfactions may motivate individuals to increase effort, or increase strategy search and development, in order to perform better; however, they may also lead to feelings of despair and inability. In order to balance the beneficial effects of increased motivation with potentially harmful effects, two guidelines have been suggested. It is recommended that goals be designed such that individuals evidence: (1) moderate goal-performance discrepancies (Bandura & Cervone, 1983, 1986), and (2) progression toward their goals (Hsee and Ableson, 1991). It is believed that when moderate goal-performance discrepancies and progression toward the goal are evidenced, individuals will conclude

that they are doing pretty well and will not devote an excessive amount of cognitive resources toward self-regulation and strategy development. In addition, it is likely that an individual's self-efficacy will be maintained or raised as a result of the improvements witnessed. In turn, performance may benefit directly or indirectly.

New Directions for Goal Assignments on Complex Tasks: The Effects of Updated and Individualized Goals. Based on the guidelines recommended above (i.e., moderate goal-performance discrepancies and goal progression), the present study will investigate two ways to design goals more effectively for complex tasks: (1) updating goals over time, and (2) tailoring goals to an individual's performance rather than the group's performance. Goal updating is a very natural process for individuals repeatedly performing an activity. Individuals may raise or lower their personal performance goals over time (Campion & Lord, 1982). Although goal updating is a natural process which occurs during personal goal setting, research on assigned goals tends to overlook the effects of goal updating. One study (Earley, Connolly, & Ekegren, 1989), however, has investigated the effects of updated goals on complex task performance.

Earley, Connolly, & Ekegren (1989) conducted a study to investigate the effects of goals on strategy development and performance on a stock market prediction task. They compared the effects of five different types of goals, including: (1) do-your-best, (2) specific-easy (come within \$30), (3) specific-moderate (come within \$20), (4) specific-hard (come within \$10), and (5) tapering, specific goals. Individuals who were given tapering (i.e., updated) goals were originally told to come within \$30 of the actual stock price. This goal was then decreased by \$5 after every 20 predictions; the final goal in this condition was to come within \$10 of the actual price. All subjects performed five blocks of a stock market prediction task, with twenty trials per block. For each trial, subjects predicted a stock price for a fictitious company and were then given feedback on the actual stock price. Overall, individuals in the do-your-best condition demonstrated better performance than individuals in the specific-difficult goal group. However, those in the specific, tapering (updated) goal condition performed nearly as well as those in the do-your-best goal condition, with performance differences decreasing over time. Subjects in the updated goal condition even outperformed those in the do-your-best goal condition in the third performance block; this difference, however, was not statistically significant. Earley et al. (1989) concluded that updating goals over time may reduce excessive strategy search that results from an overly difficult goal.

In order to create truly individualized goals which provide moderate goal-performance discrepancies for all learners, it is suggested that goals also be assigned based on each learner's progress rather than the group's progress. In psychological experiments, the level of goal difficulty assigned is typically based on a group "standard" (i.e., the 80th or 85th percentile of group performance) (Locke & Latham, 1990). So, at any point in time, all subjects receive the same goal. However, a goal which provides a moderate goal-performance discrepancy for one individual may provide a high or low discrepancy for another person. In order to provide moderate goal-performance discrepancies for everyone, goal assignments need to be tailored to each individual (rather than to the group). It is likely that goal assignments based on each individual's capabilities, preferences, and/or prior performance will lead to better complex task performance than goals based on a group standard. Since individual differences in cognitive ability, learning orientation, and task-relevant experience (e.g., prior performance) can significantly influence skill

acquisition (Ackerman & Humphreys, 1990; Kanfer & Ackerman, 1989; Weiss, 1990), it is recommended that trainers take these characteristics into consideration when assigning performance goals. Thus, truly "individualized" goals should not only tailor the goal level to each learner (to create moderate discrepancies), but also the updating of goals (i.e., when a goal is updated) to each learner. Recent developments in cognitive and instructional psychology, especially the development of Intelligent Tutoring Systems (ITSs), simplify the challenging task of diagnosing learner progress. As a result, goal individualization is a timely and viable technique for improving learning during complex task training.

The purpose of the present study is to (1) replicate past research demonstrating the detrimental effects of standard (group-based, not-updated) goals on complex task performance (in comparison to "do-your-best" goals), and (2) demonstrate that, contrary to this finding, goal setting can benefit complex task performance if goals are individualized (i.e., tailored to the individual and updated over time). It is hypothesized that, by creating moderate performance-goal discrepancies and expediting progression toward goal achievement, individualized goals will increase self-efficacy and decrease negative affect. In addition, attention to performance evaluation, off-task attention, and the amount of strategy search employed by the learner is expected to lessen. As a result, learners given individualized goals are expected to perform better than those given standard goals or do-your-best goals during skill acquisition.

Hypothesis 1: Consistent with recent findings, subjects given do-your-best goals will perform significantly better than those given difficult, standard goals.

Hypothesis 2: Subjects in the individualized goal condition will perform significantly better than subjects in the standard and do-your-best goal conditions.

Hypothesis 3: The effect of goals on performance will be mediated by self-efficacy, negative affect, off-task attention, and strategy search. More specifically, individualized goals will lead to better performance by increasing self-efficacy, and decreasing negative affect, off-task attention and attention to performance evaluation, and strategy search.

## Methodology

Subjects. Subjects were 56 individuals hired by local temporary agencies. Data was collected at Lackland AFB in San Antonio, TX. Consistent with previous ATC research (Ackerman, 1992), participation was restricted to individuals between 18 and 30 years of age (to prevent age-related decline of success in ATC skill acquisition) who did not have prior ATC experience. Although 81 subjects completed the experiment, twenty-five subjects were excluded from the analyses due to: (1) extremely low post-test scores (more than two standard deviations below the mean) (11 Ss); (2) malicious behavior (e.g., deliberately crashing planes)(2 Ss); and (3) the expressed desire to quit during the majority of the experiment (12 Ss).

The remaining 56 subjects provided the data for which subsequent analyses are reported. Means and standard deviations for individual characteristics are provided in Table 1. Seventy-eight percent of the subjects were male; ninety-eight percent had only a high school education or equivalent. The means for videogame and computer

experience were 4.00 and 3.51, respectively (on a five-point likert scale, with standard deviations close to 1); thus subjects reported having a moderate amount of videogame and computer experience, on average. The mean age was 22.8 years (s.d.= 3.3 years). Subjects were paid approximately \$5 per hour for participation. Payment was not contingent on task or test performance.

Experimental Task. The experimental task used in this experiment is a modified version of Terminal Radar Approach Control (TRACON) simulation software developed by Wesson International and used by Ackerman (1992). The task requires subjects to learn a set of rules for air traffic control, including (a) reading flight strips, (b) declarative knowledge about radar beacons, airport locations airport tower hand-off procedures, en-route center hand-off procedures, airplane separation rules and procedures, (d) monitoring strategies, and (e) strategies for sequencing airplanes for maximum efficient and safe sector traversal. This task is highly dependent on human-computer interface skills such as issuing mouse-based commands, menu retrieval, keyboard operations, and integration between visual and auditory information channels. For a more detailed description of TRACON, see Ackerman (1992).

The TRACON simulation was selected because it is a complex, real-world task. This task requires individuals to learn an extensive amount of air traffic control terminology and commands, coordinate the actions of numerous airplanes and airports in order to avoid flight delays and conflicts (e.g., crashes), and adjust to incoming planes and aircraft demands. Thus, the TRACON task meets the criteria for task complexity provided by Wood (1986). Furthermore, versions of the TRACON simulation are currently used by military and civilian schools to train future pilots and air traffic controllers on the fundamentals of air traffic control. Thus, it is felt that findings from this experiment would be highly generalizable to performance on complex tasks in real-world settings.

The version of TRACON used in this experiment was modified to allow for shortened performance trials which are typical of goal setting research. Each trial lasted 10 minutes. To allow for shorter performance trials, two critical changes were made: First, the subject's sector was reduced in terms of the size, number of airports, and number of fixes (i.e., 2 airports and 8 fixes). In addition, only 10 planes were released, of which about 5 planes could be handled successfully each ten-minute trial. Each trial contained airplanes that were divided into three basic categories (overflights, departures, and arrivals). In order to make the task trials roughly equivalent in difficulty, each trial consisted of 4 arrivals, 3 overflights, and 3 departures, which were released in the same order during each trial.

Subjects were told to perform the task so that they successfully handled as many planes as possible, keeping safety in mind. Points were given for successful accomplishment of each airplane's flight plan (ranging from about 500-1200 points), and penalty points (ranging from about 250-2000 points) were deducted for errors such as incorrect or missed hand-offs, missed approaches for arrivals, separation conflicts near misses, and crashes. Crashes, which were considered to be the most severe error, carried the steepest penalty (i.e., 2000 points). A subject's total score was shown on the screen throughout each trial. Following each trial, the assigned goal was re-iterated and performance feedback relevant to the goal (i.e., total score) was given. Following the recommendation of Ilgen and



Moore (1987) to provide feedback separately on all important performance dimensions, feedback pertaining to penalty and earned points was also provided (in addition to the total score).

**Design and Manipulations.** A three-by-three repeated measures factorial design was used to investigate the effects of goal type on performance. Subjects were assigned to one of three goal conditions: Do-Your-Best, Standard, or Individualized. Subjects were given two practice trials to familiarize themselves with the task. During practice, all subjects were told to "do your best" to maximize their total score. Following practice, goal were assigned and subjects performed three blocks of the TRACON task (with each block consisting of five 10-minute trials). Subjects in the Do-Your-Best goal condition (control group) were told to "do your best to maximize your total score" throughout the entire experiment. This group consisted of 22 subjects and served as the basis for standard and individualized goal assignments.

Standard and individualized goals were difficult, specific, and performance-oriented. In these two conditions, individuals were instructed prior to each trial to achieve a particular total score (e.g., "Your goal is to achieve a total score of 5000 on the next block of trials"), depending on their goal assignment. More specifically, subjects in the Standard goal condition were given group-based, non-updated goals. In this condition, 19 subjects were given one difficult, specific performance goal based on the 85th percentile of performance of the control group during the final performance block. Accordingly, subjects were told to obtain a total score of 2100 (unadjusted mean plus one standard deviation, or  $500 + 1600 = 2100$ ) during all three blocks of performance trials. Subjects in the Individualized goal condition were given individual-based, updated goals. In this condition, 15 subjects were assigned specific, difficult performance goals based on their predicted performance.

Using data from the do-your-best goal condition, a best-fitting regression line predicting performance was calculated for each block using several individual difference variables (e.g., cognitive ability, computer experience, learning orientation, and declarative knowledge) and task performance during the previous block. The regression equations used to predict block performance are presented in Table 2. Following each block of trials (and prior to the upcoming block), several steps were taken to obtain individualized goals. First, the computer computed the mean performance level (across the five trials) for each subject. Using this information, as well as information on the relevant individual difference variables (obtained on the first day of training), the computer then predicted each individual's upcoming performance using the appropriate regression equation. Next, an individualized goal was then calculated for each subject by incrementing the predicted performance score by one standard deviation ( $s.d.= 1600$ ). Finally, limits were placed on these individual goal assignments to prevent extremely high or low discrepancies (in case of overprediction or underprediction). Individualized goals ranged from -1000 to 2000 in the first performance block (mean=358,  $s.d.=1254$ ); from 0 to 3000 in the second performance block (mean=709,  $s.d.=1232$ ); and from 1750 to 4000 (mean=2200,  $s.d.=932$ ) in the third performance block.

**Measures.** Subjects were asked to reported demographic information (e.g., gender). In addition, subjects completed questionnaires used to assess their declarative knowledge of the TRACON task, goal orientations, and various cognitions resulting from performance of the TRACON task. Cognitive ability and task performance were also assessed.

*Demographic Questionnaire.* Consistent with previous ATC research (Ackerman, 1992), participation was restricted to individuals between 18 and 30 years of age who did not have prior ATC experience. Thus, subjects were asked to report their age and air traffic control (ATC) experience. ATC experience was assessed using a 5-point likert scale with responses ranging from (1) no experience to (5) a lot of experience. Subjects were also asked to report their gender, education, videogame experience, and computer experience. Videogame and computer experience were assessed using a 5-point likert scale with responses ranging from (1) no experience to (5) a lot of experience. Subjects were asked to report the highest level of education they have achieved: a high school diploma, GED, Associate's degree, Bachelor's degree, Master's degree, or Doctorate degree.

*Knowledge Test.* Declarative knowledge was assessed using a 25-item test developed specifically for this experiment (see Appendix 1). This test includes both true-false and multiple choice questions. Each multiple choice item has three alternatives. An example of a true-false question is "Altitude should be adjusted every time a plane is handed off": T or F. An example of a multiple-choice question is "When should a plane's speed be adjusted?" (a) every time a plane is handed off (b) every time a plane is landing (c) only when necessary to avoid a conflict. This scale had an average internal consistency of .78.

*Cognitive Ability.* Cognitive ability was assessed using the CAM 4.0 battery of computerized tests (Kyllonen, Woltz, Christal, Tirre, Shute, & Chaiken, 1990). The CAM 4.0 battery measures six different aptitudes (e.g., working memory) across three domains (quantitative, verbal, and spatial). Given the nature of the TRACON task, three aptitudes were assessed for the spatial domain: inductive reasoning skills, information processing speed, and working-memory capacity. When assessing processing speed and working memory, measures of accuracy (percent of problems correct) and completion time were obtained.

*Goal Orientation.* Two 8-item measures of learning and performance goal orientations developed by Button, Mathieu, and Zajac (1994) were used to measure goal orientation. Recent research has demonstrated that these are 2 distinguishable dimensions which are not significantly correlated. An example performance goal orientation question follows: "Performing a task better than others is important to me." An example learning goal orientation question follows: "It is important to continually improve oneself." Button et al. (1994) reported internal consistency reliabilities of .84 and .80 for the learning goal and performance goal orientation scales, respectively. Both scales use a five-point likert scale ranging from (1) Strongly Agree to (5) Strongly Disagree. Responses were recoded such that higher scale scores indicated stronger learning or performance goal orientations.

*Goal Manipulation Checks: Personal Goals and Goal Commitment.* As a manipulation check for the goal assignments, individuals were asked to report the specific score they were trying to achieve (i.e., their personal goals). Goal commitment was assessed using a eight-item scale adapted from Hollenbeck, Klein, O'Leary, and Wright's (1989) measure of goal commitment. An example question follows: "It's hard to take my assigned performance goal seriously." Subjects responded using a 5-point likert scale ranging from (1) strongly disagree to (5) strongly agree. Responses were recoded such that higher scores indicate stronger commitment to the assigned goal.

*Cognitive Measures.* Negative affect and attention allocation (off-task attention, on-task attention, and attention to performance evaluation) were assessed using items from Kanfer and Ackerman (1989) and Mikulincer (1989). An example item for the negative affect scale follows: "I thought about how dissatisfied I was with my performance on this task." An example item for the on-task attention scale follows: "I paid close attention to the areas that were difficult for me." Responses were based on a 5-point likert scale ranging from (1) never to (5) constantly, with higher scale scores indicating more negative affect or more allocation of attention. Self-efficacy was assessed using an eight-item measure adapted from Brock (1995). An example question follows: "I am confident in my ability to effectively handle many planes." Responses were based on a 5-point likert scale ranging from (1) strongly disagree to (5) strongly agree, with higher scale scores indicating higher levels of self-efficacy. Perceived performance pressure was also assessed using a four-item scale adapted from Kanfer and Ackerman (1989). An example question follows: "This task is fast-paced." Responses were based on a 5-point likert scale ranging from (1) strongly disagree to (5) strongly agree, with higher scale scores indicating more perceived pressure.

Based on the work of Earley, Connolly, and Ekegren (1989), strategy usage was assessed by asking individuals how many strategies they used to handle the planes successfully. Subjects responded by selecting one of four statements indicating frequency of strategy change. "I tried one or two methods and stayed with a single one" and "I tried several new ways to see which one was best" were considered low-search strategies, while "I kept trying new ways of handling planes and finally settled on one" and "I tried new ways to handle planes and would switch back and forth between several effective ways" were considered high-search strategies.

All goal commitment and cognitive measures were computer-administered at the end of each performance block (i.e., after every 5 trials). Table 5 displays the number of items per scale, and the estimates of internal consistency reliability for each scale (ranging from .58 to .86).

*Performance.* Overall performance for each trial was assessed using the total score achieved for the TRACON simulation. Total score was computed as the sum of all points earned (for flights handled successfully) minus any penalties for errors committed (e.g., crashes).

*Procedure.* Subjects were run in groups of 15-25 subjects, with the control group (do-your-best goal condition) run first. The experiment was completed across two days for each group of subjects, and consisted of approximately 2 hours of ability testing, 3 hours of instruction, 1 hour of review and knowledge testing, and 3 hours of TRACON task performance. During the morning of the first day, subjects participated in individual difference testing (i.e., demographic questionnaire, CAM aptitude tests, and goal orientation measure). The afternoon of the first day was then devoted to training subjects how to perform the TRACON simulation.

First, subjects watched a videotape which taught key terms and commands needed to perform TRACON. The videotape was developed by Ackerman (1992) to describe the major task components, the rules regarding operation of the computer interface, display characteristics, and procedures for accomplishing the controller task. The tape (110 minutes in length) contained three main sections: overflights, departures, and arrivals. Subjects alternated between the videotape viewing and hands-on practice. After subjects learned about each type of flight, they completed a part-task training trial in which they practiced handling two flights. Following the three part-task

training sessions, an instructor briefly reviewed the TRACON task using a short summary manual, and answered subjects' questions.

A 25-question declarative knowledge test was then administered via the computer. This test was considered to be part of the training; thus following test administration, all questions were reviewed via the computer. Correct responses were highlighted on the computer (as well as the subjects' responses, if incorrect) and reviewed by the instructor. Lastly, subjects were given two practice trials to familiarize themselves with the full TRACON task. The practice trials were created identical to the performance trials. Trainees were allowed to use their manuals during the practice trials (but not during the performance trials). During the practice trials, all subjects, regardless of goal assignment, were told to "do your best" to maximize the total score. Following the practice trials, the goal commitment and cognitive questionnaires were administered. This information provided an assessment of performance, goal commitment, and cognitions at baseline (i.e., prior to the goal manipulations).

On day 2, subjects were asked to re-read their TRACON manuals. The instructor then gave a brief review of TRACON and administered the declarative knowledge test once again. The purpose of the second testing was to (1) assess test improvement and learning retention, and (2) ensure that subjects were trained sufficiently to perform the TRACON simulation. Subjects who obtained a score two or more standard deviations below the mean (18 or below) completed the experiment, but were excluded from the analyses. Following this testing, the performance trials began. Each performance block consisted of five 10-minute trials followed by a self-report questionnaire. Short breaks were given after each performance block. Subjects were told not to discuss the task with each other during the breaks. At the conclusion of the experiment, subjects were debriefed and paid.

## Results

For all analyses, responses across the 15 performance trials were blocked into three groups of 5 trials. Means, standard deviations, and intercorrelations for the individual difference measures and performance are provided in Table 1. To ensure group equivalence prior to the goal manipulation, means for post-training knowledge, goal commitment, cognitive ability, and practice performance were compared across the three conditions. Several one-way ANOVAs were performed, using goal type as the independent variable. Findings indicated no significant differences for declarative knowledge and goal commitment, but significant differences for cognitive ability and practice performance. Thus, cognitive ability and practice performance were covaried in subsequent analyses. Since the spatial ability test for inductive reasoning demonstrated the strongest correlations with performance, inductive reasoning scores were used when controlling for cognitive ability in subsequent analyses.

To assess the efficacy of the goal manipulations, performance-goal discrepancies were examined across goal conditions. Since subjects in the do-your-best goal condition did not receive an assigned goal, discrepancies between performance and assigned goal could not be evaluated. Based on the goal manipulation, it was expected that the performance-goal discrepancies for the standard goal group should decrease over time as subjects progressed toward the non-updated goal, while discrepancies for the individualized goal group should remain moderate and

nearly equivalent across time. Results indicate that for the standard goal group, mean discrepancies decreased across the three performance blocks from 3466, to 2342, to 1975. Furthermore, mean performance-goal discrepancies for the individualized group increased from blocks 1 to 2, but leveled off during block 3 (mean discrepancies equal to 1853, 2522, and 2478, respectively). Given that the mean performance-goal discrepancies for the individualized goal group were higher than those for the standard goal group during the last two blocks of performance, the efficacy of the goal manipulations is equivocal. To investigate the effects of goal assignment and block on goal commitment and personal goals, 3 x 3 (goal type x block) repeated measures analyses of variance (ANOVAs) were conducted. Goal commitment did not differ across blocks (see Table 5) or goal conditions. Furthermore, goal setting did not have a significant effect on the personal goal that participants reported adopting ( $F(2,46)=1.51, p>.05$ ); however, personal goals did increase significantly across blocks ( $F(2,92)=3.83, p<.05$ ).

To test for effects of goal type on performance (H1 and H2), a 3 X 3 repeated measures ANCOVA was performed on overall task performance (i.e., total score). The independent variables were goal type and block; covariates used include cognitive ability (inductive reasoning) and practice performance. Adjusted means are shown in Table 3. The results of the repeated measures ANOVA are presented in Table 4. There was a significant effect for block, with performance increasing over time. In addition, a significant goal x block interaction was found. To determine the locus of the interaction, a test of simple main effects was used within each trial block. Fisher protected *t* tests indicated that for Block 1, subjects in the do-your-best condition significantly outperformed those in the standard goal condition. Furthermore, in Block 2, subjects in the do-your-best and standard goal conditions significantly outperformed subjects in the individualized goal condition. No significant differences in performance were found for Block 3. Although the differences were only significant during the first block, subjects in the do-your-best condition outperformed subjects in the standard goal condition during all three performance blocks. Thus, the data provide partial support for hypothesis 1. Individualized goal subjects outperformed standard goal subjects during block 1; however, this effect was not statistically significant. Furthermore, they performed worse than subjects in the do-your-best and standard goal conditions during blocks 2 and 3 of performance (with a noticeable drop in performance occurring between blocks 1 and 2).

Means, standard deviations, and intercorrelations for the cognitive measures (negative affect, self-efficacy, pressure, and attention allocation) are shown in Table 5. The univariate statistics are presented for each block. To determine whether goal type or task experience (block) affected these cognitions, 3x3 repeated measure ANOVAs were conducted. Goal type and block were the independent variables. Negative affect, self-efficacy, pressure, and attention (on-task cognitions, off-task cognitions, and performance evaluation) were the dependent variables. All cognitions differed significantly across time, but not across goal conditions. Self-efficacy and on-task cognitions increased over time, while task pressure ( $F_{(2,106)}=7.69, p<.01$ ), negative affect ( $F_{(2,116)}=8.03, p<.01$ ), performance evaluation ( $F_{(2,106)}=5.99, p<.01$ ), and off-task cognitions ( $F_{(2,106)}=9.67, p<.01$ ) decreased across blocks.

To investigate the effects of task strategy, responses to the strategy usage item were dichotomized into low-search and high-search strategies. "I tried one or two methods and stayed with a single one" and "I tried several new ways to see which one was best" were considered low-search strategies, while "I kept trying new ways of handling

lanes and finally settled on one" and "I tried new ways to handle planes and would switch back and forth between several effective ways" were considered high-search strategies. The percentage of individuals in each condition who reported frequent strategy switching (high-search strategies) are shown in Table 6. A chi-square test indicated no significant differences across conditions and performance blocks. Although not statistically significant, subjects in the individualized condition consistently reported more strategy switching than subjects in the other two conditions. Given that the goal manipulation had only marginal effects on performance and no significant effects on cognitions, a formal test of the mediation effects is not warranted. Thus, hypothesis 3 was not evaluated.

### Conclusion

The results indicate that individuals given do-your-best goals consistently outperformed individuals given standard goals. This effect, however, was only statistically significant during the first block of performance; thus, hypothesis 1 was partially supported. Consistent with past research, the assignment of group-based, fixed goals on complex or novel tasks was detrimental to performance. Contrary to hypothesis 2, subjects given goals which tailored to one's aptitudes and performance levels (i.e., individualized goals) did not consistently outperform subjects given do-your-best or standard goals. Subjects given individualized goals did outperform standard goal subjects during the first block, thus providing initial support for hypothesis 2. However, this performance advantage disappeared during blocks 2 and 3. In fact, individualized goal subjects actually exhibited a drop in performance during the second block (the only performance drop which occurred for any of the groups). Given that goal type did not have a significant main effect on performance, formal mediation tests were not warranted and hypothesis 3 was not evaluated.

So why did individualized goal subjects initially outperform standard goal subjects, while performing significantly worse during the second block? Several explanations are possible. First, the average performance-goal discrepancies for the individualized goal group increased from block 1 to block 2. Furthermore, the average performance-goal discrepancies for the individualized goal condition were larger than those experienced by the standard goal group during blocks 2 and 3. Thus, it appears that individualized goals were not designed appropriately such that goal-performance discrepancies were moderate. This pattern of discrepancies seems to be consistent with the pattern of strategy switching reported. Subjects in the individualized condition consistently reported more strategy switching than subjects in the other two conditions. This effect was most notable during the second performance block. In the individualized goal condition, 40% of subjects reported high-search strategies during the first performance block, while 55% reported high-search strategies during the second block; thus, there was an increase in the number of individualized goal subjects who reported frequent strategy switching following block 1. In addition to an increase in strategy search within condition, more individualized goal subjects (55%) reported frequent strategy switching than subjects given standard goals (32%) or do-your-best goals (27%). This increase in strategy switching corresponds to the drop in performance demonstrated by the individualized goal subjects during the second performance block, and is consistent with previous findings indicating that more strategy switching tends to debilitate performance (Earley, Connolly, & Lee, 1989).

Although the hypotheses were not fully supported, the data do indicate trends in the expected directions for hypothesis 1, as well as some support for hypothesis 2 during the first block. It is likely that the lack of statistical support for hypothesis 1 was due to low power as a result of the small sample used. Furthermore, it is felt that lack of support for hypothesis 2 was due to methodological problems stemming from the design of individualized goals. Future research should attempt to replicate these findings before it can be concluded that individualized goals do not benefit complex task performance. Given that jobs have become increasingly more novel and complex, interventions which will motivate individuals when learning and performing complex tasks are now more important than ever. It is felt that providing individualized goals is one way to increase motivation on complex tasks, and thereby, learning and performance. Although several recommendations for goal individualization have been made based on previous research and theory (i.e., providing moderate goal-performance discrepancies and goal progression), many more questions need to be answered before assigned goals can be truly effective on complex tasks. For example, what individual difference variables should be considered when tailoring goals to individuals? How often, and by how much, should goals be updated? Following poor performance, should goals still be updated? If so, should they be elevated or reduced? Although individualized goals were not practical in the past due to the labor-intensive process of continuously evaluating a learner's progress and updating his or her goal, recent developments in Intelligent Tutoring Systems (ITSs) have made the endeavor of individualized goals possible.



## References

- Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. Journal of Applied Psychology, 77, 598-614.
- Ackerman, P. L., & Humphreys, L. G. (1990). Individual differences theory in industrial and organizational psychology. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of Industrial and Organizational Psychology (2nd Ed.), Vol. 1. Palo Alto, CA: Consulting Psychologists Press.
- Bandura, A. (1982). The self and mechanisms of agency. In J. Suls (Ed.), Psychological perspectives on the self (Vol. 1, pp. 3-39). Hillsdale, NJ: Erlbaum.
- Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. Journal of Personality and Social Psychology, 45, 1017-1028.
- Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. Organizational Behavior and Human Decision Processes, 38, 92-113.
- Brock, D. B. (1995). Measuring Self-Efficacy: Conceptual, Computational, and Practical Problems. Paper presented at the 10th Annual Conference of the Society for Industrial and Organizational Psychology: Orlando, FL.
- Button, S.B., Mathieu, J.E. & Zajac, D.M. (1994). The Development and psychometric evaluation of measures of learning goal and performance goal orientation. Manuscript under review, The Pennsylvania State University.
- Campion, M. A., & Lord, R. G. (1982). A control systems conceptualization of the goal-setting and changing process. Organizational Behavior and Human Performance, 30, 265-287.
- Carver, C. S., & Scheier, M. F. (1981). Attention and self-regulation: A control theory approach to human behavior. New York: Springer-Verlag.
- Dweck, C. S. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.
- Earley, P. C., Connolly, T., & Ekegren (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. Journal of Applied Psychology, 74, 24-33.
- Erez, M. (1977). Feedback: A necessary condition for the goal setting-performance relationship. Journal of Applied Psychology, 62, 624-627.
- Goldstein, I., & Gilliam, P. (1990). Training system issues in the year 2000. American Psychologist, 45, 134-143.
- Hollenbeck, J. R., Klein, H. J., O'Leary, A. M., & Wright, P. M. (1989). Investigation of the construct validity of a self-report measure of goal commitment. Journal of Applied Psychology, 74, 951-956.
- Howell, W. C., & Cook, N. J. (1989). Training the human information processor. In I. Goldstein (Ed.), Training and development in work organizations (pp. 121-182). San Francisco: Jossey-Bass.
- Hsee, C.K., & Ableson, R.P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. Journal of Personality and Social Psychology, 60, 341-347.
- Johnson, D. S., & Kanfer, R. (1992). Goal-performance relations: The effects of initial task complexity and task practice. Motivation and Emotion, 16(2), 117-141.



- Ilgen, D. R., & Moore, C. F. (1987). Types and choices of performance feedback. Journal of Applied Psychology, 72, 401-406.
- Kyllonen, P.C., Woltz, D.J., Christal, R.E., Tirre, W.C., Shute, V.J., & Chaiken, S. (1990). CAM 4.0: Computerized Battery of Cognitive Ability Tests. Unpublished computer program, Brooks Air Force Base, TX.
- Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of Industrial and Organizational Psychology (2nd Ed.), Vol. 1 (pp. 75-170). Palo Alto, CA: Consulting Psychologists Press.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integration/aptitude-treatment interaction approach to skill acquisition. Journal of Applied Psychology, 74, 657-690.
- Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. Organizational Behavior and Human Decision Processes, 50, 212-247.
- Locke, E.A., & Latham, G.P. (1990). A theory of goal setting & task performance. Englewood Cliffs, NJ: Prentice-Hall.
- Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. Psychological Bulletin, 90, 125-152.
- Matsui, T., Okada, A., & Inoshita, O. (1983). Mechanism of feedback affecting task performance. Organizational Behavior and Human Performance, 31, 114-122.
- Mento, A. J., Steel, R. P., & Karren, F. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966-1984. Organizational Behavior and Human Decision Processes, 39, 52-83.
- Mikulincer, M. (1989). Cognitive interference and learned helplessness: The effects of off-task cognitions on performance following unsolvable problems. Journal of Personality and Social Psychology, 57, 129-135.
- Smith, K. G., Locke, E. A., & Barry, D. (1990). Goal setting, planning, and organizational performance: An experimental simulation. Organizational Behavior and Human Decision Processes, 46, 118-134.
- Weiss, H. M. (1990). Learning theory and industrial and organizational psychology. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of Industrial and Organizational Psychology (2nd Ed.), Vol. 1 (pp. 171-221). Palo Alto, CA: Consulting Psychologists Press.
- Wood, R. E. (1986). Task complexity: Definition of the construct. Organizational Behavior and Human Decision Processes, 37, 60-82.
- Wood, R. E., Bandura, A., & Bailey, T. (1990). Mechanisms governing organizational productivity in complex decision-making environments. Organizational Behavior and Human Decision Processes, 46, 181-201.
- Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. Journal of Applied Psychology, 72, 416-425.

**Table 1****Means, Standard Deviations, and Intercorrelations of Overall Performance and Individual Characteristics (n=56)**

Variable	Mean	S.D.	Correlations <sup>a</sup>			
			Practice Performance	Block 1 Performance	Block 2 Performance	Block 3 Performance
1. Practice Performance	-3042.82	2493.71				
2. Block 1 Performance	-1039.61	2035.04	.41**			
3. Block 2 Performance	-442.04	2086.57	.35**	.73**		
4. Block 3 Performance	134.04	1884.99	.44**	.66**	.76**	
5. Age	22.79	3.28	-.02	-.07	.02	-.05
6. Sex	.23	.43	-.22	-.13	-.09	-.11
7. Education	1.16	.42	.19	-.10	-.05	-.07
8. Video Game Experience	4.00	.96	-.02	.20	.16	.17
9. Computer Experience	3.51	.98	-.06	.25*	.29*	.28*
10. Learning Orientation	1.84	.47	.03	-.12	-.06	.06
11. Performance Orientation	2.19	.53	.00	.03	.07	.04
12. Processing Speed Accuracy	85.03	19.59	.16	.24*	.30*	.23*
13. Processing Speed Time	245.04	85.03	-.01	-.30*	-.27*	-.27*
14. Working Memory Accuracy	49.61	22.84	.19	.36**	.36**	.35**
15. Working Memory Time	925.85	105.22	-.28	-.47**	-.37**	-.41**
16. Inductive Reasoning	57.82	21.73	.15	.51**	.47**	.40**

<sup>a</sup> \* indicates  $p \leq .05$  (one-tail test); \*\* indicates  $p \leq .01$  (one-tail test)

**Table 2****Regressions Predicting Block Performance for Individualized Goals**

Predicted Block 1 Performance = 1240.82 (Computer Experience) - 7.15 (Working Memory Time)  
- 9.19 (Processing Speed Time) + 3486.01

Predicted Block 2 Performance = .60 (Block 1 Performance) - 652.33 (Learning Orientation)  
+ 103.54 (Knowledge Pre-test) + 11.49 (Working Memory Time)  
- 13.42 (Processing Speed Time) - 7385.61

Predicted Block 3 Performance = .85 (Block 2 Performance) + 1137.22 (Learning Orientation)  
+ 149.24 (Knowledge Pre-test) - 5126.95

**Table 3**

**Adjusted Means for Overall Performance Across Blocks and Experimental Conditions**

Condition	Block1	Block2	Block3
Do Your Best	-557.42	288.84	324.45
Standard	-1586.23	-383.17	86.43
Individualized	-951.77	-1410.07	43.43

Note: Means are based on 53 observations and have been adjusted for inductive reasoning and practice performance

**Table 4**

**Repeated Measures Analysis of Variance Results for Overall Performance**

Variable	MS	F	Eta <sup>2</sup>
Covariates <sup>a</sup> (C)	93890277.0	16.20**	.40
Goal Setting (G)	9696568.2	1.67	.07
Blocks (repeated)	17992634.0	16.26**	.25
G x Blocks	3523381.0	3.18*	.11

<sup>a</sup>Covariates included inductive reasoning and practice performance.

<sup>b</sup>*dfs* are (2,48) for Covariates and Goal Setting; (2,100) for Blocks; and (4,100) for Goal X Blocks.

\* indicates  $p \leq .05$ ; \*\* indicates  $p \leq .01$

**Table 5**  
Means, Standard Deviations, and Intercorrelations for Averaged Measures (n=56)

Scale	Means and Standard Deviations			Correlations <sup>a</sup>						
	Block 1	Block 2	Block 3	1	2	3	4	5	6	7
1. Goal Commitment	3.69 (.37)	3.68 (.48)	3.74 (.49)	.78(8) <sup>b</sup>						
2. Negative Affect	2.53 (.58)	2.54 (.70)	2.42 (.64)	-.26*	.78 (8)					
3. Self-Efficacy	3.50 (.58)	3.52 (.55)	3.64 (.51)	.53**	-.45**	.68 (8)				
4. Pressure	3.58 (.53)	3.52 (.57)	3.34 (.68)	-.01	.41**	-.39**	.61 (4)			
5. Perf. Evaluation	2.53 (.89)	2.50 (.91)	2.49 (1.06)	-.06	.23*	-.04	.03	.86 (2)		
6. Off-Task Cognitions	1.52 (.61)	1.52 (.61)	1.48 (.62)	-.45**	.08	-.24*	-.05	.21	.79 (5)	
7. On-Task Cognitions	3.54 (.63)	3.54 (.80)	3.71 (.74)	.47**	.12	.21	.12	.12	-.36**	.58 (6)

<sup>a</sup>scale correlations are based on the mean of scale responses across blocks; \* indicates  $p \leq .05$  (one-tail test), \*\* indicates  $p \leq .01$  (one-tail test)  
<sup>b</sup>alpha and number of items for each scale are shown in the diagonal

**Table 6**  
Percentage of individuals by block and goal condition who reported frequent strategy switching (high strategy search)

	Practice	Block 1	Block 2	Block 3
Do Your Best	41%	22%	27%	41%
Standard	42%	26%	32%	37%
Individualized	40%	40%	53%	47%

THE OCCURRENCE OF APOPTOSIS FOLLOWING ACUTE  
EXPOSURE TO TRICHLOROETHYLENE IN THE LIVERS  
OF 12-WEEK OLD MALE B6C3F1 MICE

Matthew C. Grothaus  
Graduate Student  
Department of Pathology

The Medical College of Ohio  
3000 Arlington Ave  
Toledo, OH 43614

Final Report for:  
Summer Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995

THE OCCURRENCE OF APOPTOSIS FOLLOWING ACUTE  
EXPOSURE TO TRICHLOROETHYLENE IN THE LIVERS  
OF 12-WEEK OLD MALE B6C3F1 MICE

Matthew C. Grothaus  
Graduate Student  
Department of Pathology  
The Medical College of Ohio

Abstract

Trichloroethylene (TCE) is a degreasing solvent that has been widely used by the U.S. Air Force which has resulted in groundwater contamination at and around some Air Force bases (4). TCE has been shown to be a liver carcinogen in mice (6,7). The potential for carcinogenicity of TCE in humans is largely unknown. Understanding the mechanism of liver tumorigenicity in mice is critical in assessing the cancer health risk of TCE-exposure in humans. Cell proliferation is thought to be an important early event in chemically-induced carcinogenicity. Apoptosis, a physiological form of cell death that acts as a counterpart to mitosis, is thought to function in concert with cell proliferation to maintain tissue homeostasis. A decreased incidence of apoptosis would disrupt tissue homeostasis thereby favoring cell proliferation. This has been proposed as one mechanism of clonal expansion of pre-neoplastic and neoplastic cells (5). The purposes of this study, therefore, were to determine the effects that exposure to trichloroethylene has on the incidence of apoptosis in the livers of 12-week old male B6C3F1 mice, and investigate the relationship that apoptosis has to cell proliferation that is known to occur at early time points following acute exposure to this chemical. Twelve-week old B6C3F1 mice were administered TCE at dose levels of 1200 mg/kg/day and tissue samples were harvested at days 2, 10, 14 and 28. *In situ* end labeling was used to evaluate apoptosis. Results were not significant suggesting that there are other mechanisms playing a more significant role in the formation of proliferative lesions that are present following continuous exposure to TCE.

**THE OCCURRENCE OF APOPTOSIS FOLLOWING ACUTE  
EXPOSURE TO TRICHLOROETHYLENE IN THE LIVERS  
OF 12-WEEK OLD MALE B6C3F1 MICE**

**Matthew C. Grothaus**

**INTRODUCTION**

Trichloroethylene ( TCE ) is a highly volatile liquid that is widely used as a degreasing agent for fabricated metals. The widespread use of TCE by the U.S. Air Force as a metal degreaser and cold parts cleaner has resulted in contamination at and around military bases (4). Environmental contamination has caused the U.S. Air Force to spend millions of dollars a year in the control and cleanup of TCE. The National Toxicology Program (NTP) has shown TCE to be a mouse liver carcinogen when its administration occurs by corn oil gavage at high dose levels (6,7). The carcinogenic effect of TCE exposure to humans is largely unknown. These facts coupled with the widespread contamination of groundwater has caused considerable debate regarding the potential of TCE to adversely effect human health.

Apoptosis is a physiological form of cell death that is regulated and occurs during embryonic development, immune system maturation, normal tissue homeostasis and oncogenesis. Apoptosis is considered to be an active process that may have several pathways leading to it (3). Kerr et al. ( 1972 ) have suggested that apoptosis is a normal physiological condition that serves to regulate the size of overall cell populations (5). Apoptosis is also known to occur during the regression of liver hyperplasia following the withdrawal of a promoting agent (2). Tissue homeostasis is thought to be a molecular signal-

regulated balance between the active processes of cell proliferation and physiological cell death (apoptosis).

TCE has been shown to cause oxidative stress by the generation of free radicals in B6C3F1 mouse liver slices (8) and lipid peroxidation *in vivo* (1). Oxidative stress can affect tissue homeostasis by altering both molecular signal transduction and genomic expression which could result in inappropriate cell growth and/or modulation of apoptosis, the counterpart to cell division by mitosis.

In a 60-day subchronic exposure to trichloroethylene via gavage in corn oil, male B6C3F1 mice exhibited a mitotic response at days 2-10, with maximum levels of cell proliferation occurring around day 10 and the return to normal background levels by day 14 (1). The detection of any differences in the amount of apoptosis in hepatocytes between TCE-exposed and control mice following acute exposures to trichloroethylene could aid in the understanding of the relationship between this type of active cell death and cell proliferation. This may also help to explain the effects that this chemical has on the liver early in the carcinogenic process. The purposes of this study, therefore, were to determine the effects that exposure to trichloroethylene via gavage in corn oil has on the incidence of apoptosis in the livers of 12 -week old male B6C3F1 mice, and investigate the relationship that apoptosis has to cell proliferation that is known to occur at early time points following acute exposure to this chemical.



## METHODOLOGY

### Animal Handling

Male B6C3F1 mice were obtained at 12-weeks of age from Charles River Laboratories ( Kingston, NY ). The mice were housed 5 per plastic cage and maintained on a 12-hour light/dark cycle at a constant temperature of  $22 \pm 1^{\circ}$  C with 35-50 % humidity. The animals were fed Purina Formulab # 5008 and ad lib water was provided. One week prior to the start of the study, magnetic identification micro-transponders ( BMDS, Biomedic Data Systems, Maywood, NJ ) were implanted subcutaneously according to the manufacturer's instructions. Mice were randomly assigned to treatment groups and then housed in the appropriate cages. All animals were housed in the same room under the previously stated conditions. Animal weights were recorded weekly beginning on day 1.

Trichloroethylene, 99.5 % with anti-oxidant additives, was purchased from Aldrich Chemical Company( Milwaukee, Wi., Lot No. MF 01428 EF ). Mazola<sup>™</sup> corn oil ( Best Foods, Somerset, NJ ) was used to prepare gavage solutions of trichloroethylene. The solutions were concentrated to 0 and 1200 mg/kg body weight trichloroethylene. Control groups in the study included a vehicle control group and a water only group. Trichloroethylene gavage solutions were prepared fresh each week and the concentrations were adjusted according to the mean body weight of each treatment group. Gas chromatography was employed following preparation and at weeks end to insure that the solution maintained potency. Gavage solutions were stored at  $4^{\circ}$  C. Gavage was performed at the same time each day, five days a week for eight weeks. Close observation occurred after dosing.

### Tissue Harvest / Sample Preparation

Tissue samples were taken on days 2, 10, 14 and 28 upon dosing and 5 mice from each treatment and control group were evaluated for each of these days. After euthanasia by CO<sub>2</sub>, tissue harvest took place at exactly 1300 hours on each of these days to control for time between dosing and harvest. Livers were excised and a 2 mm- thick cross section of the entire median lobe was rapidly placed in 10 % neutral-buffered formalin for 15 hours, transferred to 70% ethanol, processed and embedded in paraffin, and 3-4 micron sections cut and placed on ChemMate™ glass slides (Biotech, Santa Barbara, CA).

### In Situ Apoptosis Detection

Apoptosis was detected using the Apoptag™ In Situ apoptosis detection kit with a peroxidase labeling system (ONCOR, Gathersberg, MD). This procedure was automated using the Biotech Tech Mate 1000™ (Biotech, Santa Barbara, CA) which utilizes microcapillary gap technology. This automated procedure allowed all the specimens to be processed and stained at one time utilizing batch prepared reagents which minimized procedural variability that is inherent in the manual method.

Following deparaffinization in xylene and graded ethanols, slides were rinsed in distilled / deionized water and incubated with a diluted trypsin solution ( 1: 250 with enzyme diluent ) at 37 ° C for 15 minutes. Slides were then rinsed in 4 changes of distilled water and endogenous peroxidase was quenched by placing the slides in 3 % hydrogen peroxide in phosphate-buffered saline ( PBS ) for 6 minutes. Following exposure to an equilibration buffer ( ONCOR, S7100-1, Gathersberg, MD ) for 1 minute, sections were incubated at 37 ° C for 1 hour with working-strength terminal deoxynucleotidyl transferase ( ONCOR, S7100-3, Gathersberg, MD ) and reaction buffer ( ONCOR, S7100-2,

Gathersberg, MD ) at a ratio of 1:2, respectively. The terminal deoxynucleotidyl transferase is used to catalyze the reaction by which digoxigenin-nucleotides are incorporated into the 3'-OH ends of double- or single-stranded DNA breaks. Following this incubation, slides were again incubated with a working-strength stop / wash buffer ( ONCOR, S7100-4, Gathersberg, MD) at 37 ° C for thirty minutes. A peroxidase-linked anti-digoxigenin antibody with a removed Fc portion ( ONCOR, S7100-5, Gathersberg, MD ) was then applied to tissue sections. This allowed for optimal labeling of new 3'-OH double-stranded DNA breaks that are known to occur in high concentrations during apoptosis. After 3 rinses in PBS, the chromagen diaminobenzidine ( DAB ) was added to the sections for 9 minutes. The DAB is known to react with the bound peroxidase that is present on the anti-digoxigenin antibody thus enabling localization of the 3'-OH double stranded DNA breaks. Tissue sections were counterstained in hematoxylin ( Biotech, Santa Barbarbara, CA ) for 1 minute, dehydrated to xylene and coverslips were mounted using Permount™.

Estrus-phase uterine tissue was used as a positive control and apoptosis was quantitated based on the number of positively labeled cells per 10 mm<sup>2</sup> of tissue in combination with the morphological criteria for apoptosis (5). A negative control section was made by substituting water for the terminal deoxynucleotidyl transferase in the working-strength solution.

#### Statistical Analysis

Data was plotted and treatment differences were determined by ANOVA and Student's t-test using SigmaStat™ and SigmaPlot™ software.

#### RESULTS

### Body Weight / Gross Pathology

Weight gain was not effected by TCE dosing over the time course of the study ( Figure 1 ). At necropsy, TCE- or corn oil-treated animals did not appear to have increased body fat when compared to the water control group.

### In Situ Detection of Apoptosis

There was a low incidence of apoptosis detected in all groups and no significant treatment-related differences in the amount of apoptotic hepatocytes between treatment groups and controls were found ( Figure 2 ). The 14-day corn oil control group showed the highest incidence of apoptosis of all groups in the study ( mean =  $2.89 \pm 1.89$  apoptotic hepatocytes /  $10 \text{ mm}^2$  of tissue ). In TCE-treated mice, the 2-day exposure group exhibited the highest incidence of apoptosis ( mean =  $2.13 \pm 1.38$  apoptotic hepatocytes /  $10 \text{ mm}^2$  of tissue ). At day 28, both treatment and control groups showed a significant decrease in hepatocytes labeled for apoptosis when compared to all other days.

### DISCUSSION

One plausible mechanism of TCE-induced liver tumorigenesis in B6C3F1 mice is the clonal expansion of pre-existing initiated hepatocytes under the potential promoter influence of this chemical. Because apoptosis is thought to be the counterpart to cell division by mitosis, downward regulation of programmed cell death by a potential apogen such as TCE would favor cell proliferation and clonal expansion.

Trichloroethylene has been shown to cause an increase in cell proliferation at 2-10 days following administration of TCE via gavage in corn oil (1). It has been suggested that the inhibition of apoptosis may be a

potential mechanism by which clonally-expanding cells progress to neoplasia (5).

The results of this study indicate that exposure to TCE via gavage in corn oil does not significantly affect the incidence of hepatocyte apoptosis in 12-week old B6C3F1 mice, suggesting that there are other mechanisms playing a more significant role in the formation of the proliferative lesions that are present following continuous exposure to TCE.

An apparent limitation to this study was the inherent low incidence of apoptotic hepatocytes detected in the control mice, providing the base line for comparison. The fact that these young mice were still in an active growth phase ( see Figure 1 ), presumably still adding some liver mass, which would favor cell proliferation not apoptosis, may have contributed to the low incidence detected.

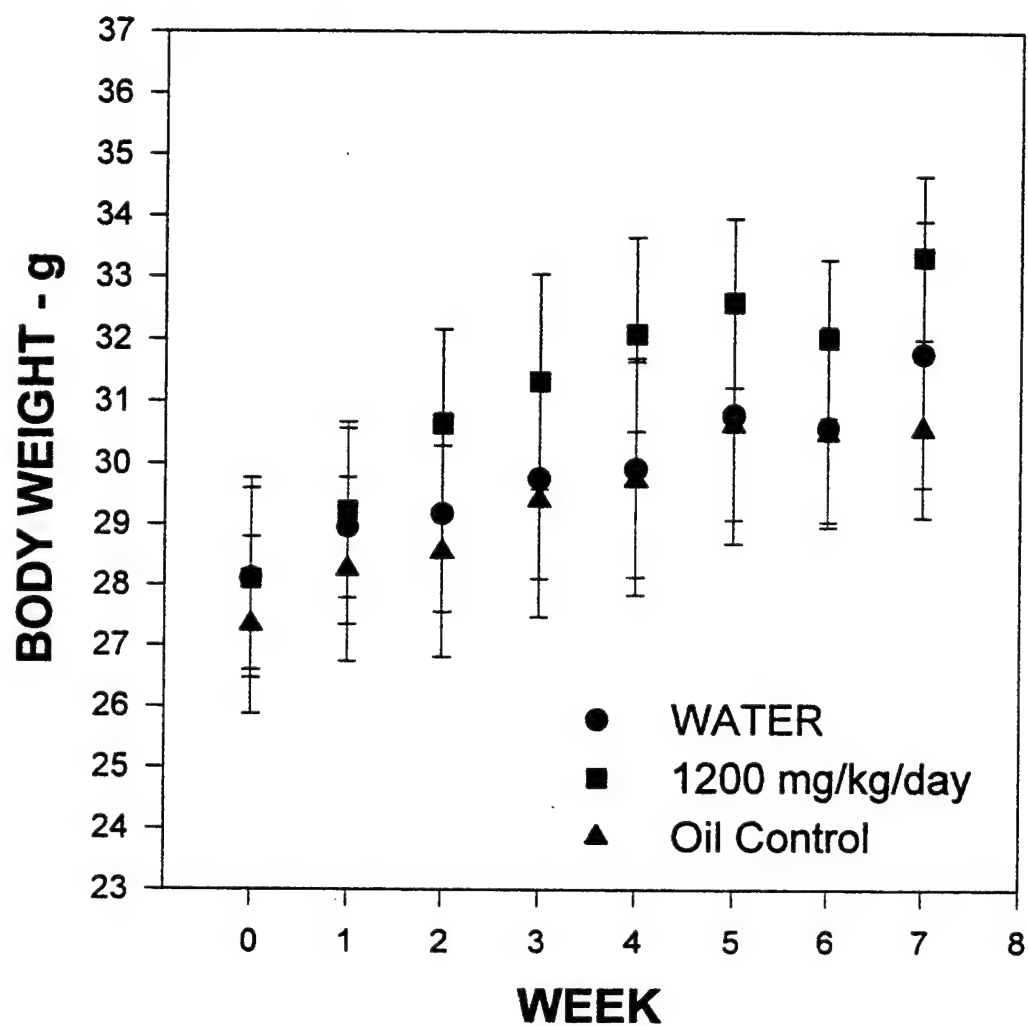
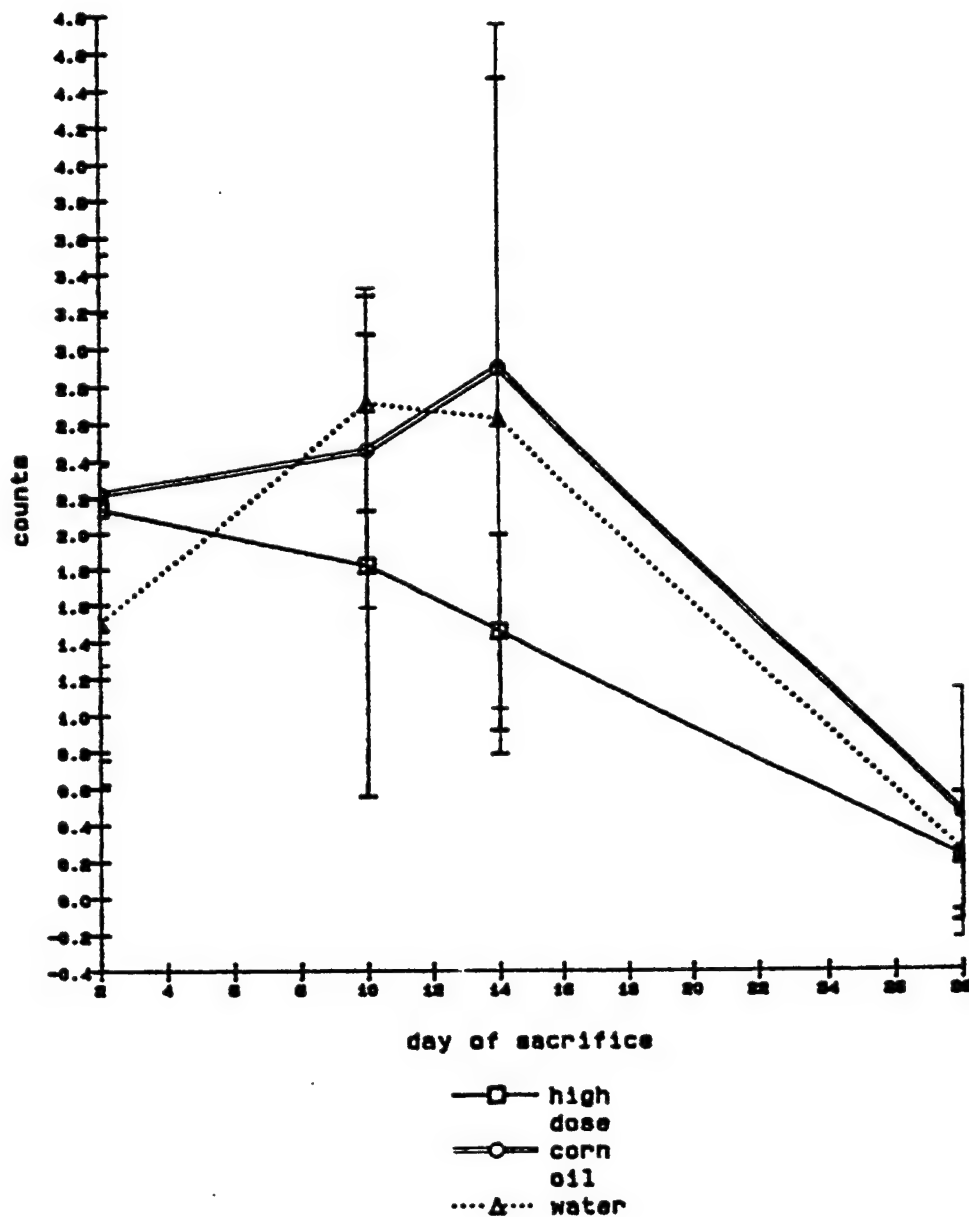


FIGURE 1. Body Weight Time Course.

Weight gain over the 8 week study was not significantly affected by TCE gavage at dose levels of 1200 mg / kg / day.

Figure 2: Incidence of apoptosis in 12-week old B6C3F1 mice<sup>a</sup> following exposure to TCE.



\*Counts represents the number of apoptotic cells found per 10 mm<sup>2</sup> of tissue.

<sup>a</sup> N=5 for treatment and control groups.

## References

1. Channel, S.R. ( 1995 ) A sub-chronic exposure to trichloroethylene causes lipid peroxidation and cell proliferation in male B6C3F1 mice. Unpublished results.
2. Columbano, A., Ledda-Columbano, G.M., Coni, P., Faa, G., Liguori, C., Santacruz, G. and Pani, P. (1985) Occurrence of Cell Death (Apoptosis) during Involution of Liver Hyperplasia. Lab. Invest. 52: 670.
3. Ellis, R.E., Yuan, J. and Horvitz, H.R. ( 1991 ) Mechanisms and Functions of Cell Death. Annu. Rev. Cell. Biol. 7: 663-98.
4. IARC (International Agency for Research on Cancer, Trichloroethylene; 1979), IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, 20, Lyon France, pp. 545-572.
5. Kerr, J.F.R., Wyllie, A.H. and Currie, A.R. ( 1972 ) Apoptosis: A Basic Biological Phenomenon with Wide Ranging Implications in Tissue Kinetics. Br. J. Cancer 26: 239-57.
6. NCI (1976) Carcinogenesis Bioassay of Trichloroethylene. National Cancer Institute, Department of Health, Education and Welfare, Bethesda, MD. NCI-CG-TR-2.
7. NTP (1990) Carcinogenesis studies of trichloroethylene (without epichlorhydrin) (CAS No. 79-01-6) in Fisher-344/N rats and B6C3F1 mice (gavage studies). National Toxicology Program Technical Report No. 243. Research Triangle Park, NC.
8. Steele-Goodwin, L., Pravecsek, T.L., Hancock, B.L., Schmidt, W.J., Channel, S.R., Bartholomew, D., Bishop, C.T. and Ketcha, M.M. (1995) Trichloroethylene: Free Radical Studies in B6C3F1 Mouse Liver Slices. Unpublished results.



**REDUCTIVE DEGRADATION AND SORPTION OF  
*cis-* AND *trans*-1,2-DICHLOROETHENE IN A METALLIC  
IRON/WATER SYSTEM**

Richelle M. Allen-King, Assistant Professor and  
Richard M. Halket, Graduate Student

Department of Geology  
Washington State University  
Pullman, WA 99164-2812

Final Report for:  
Summer Faculty Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1995

# REDUCTIVE DEGRADATION AND SORPTION OF *cis*- AND *trans*-1,2-DICHLOROETHENE IN A METALLIC IRON/WATER SYSTEM

Richelle M. Allen-King, Assistant Professor and  
Richard M. Halket, Graduate Student  
Department of Geology  
Washington State University  
Pullman, WA 99164-2812

## ABSTRACT

Reductive transformation kinetic and sorption coefficients were determined for both *cis*- and *trans*-1,2-dichloroethene (DCE) in batch systems with zero-valent iron and water. Chloride was produced by the transformation reaction and chlorine mass balances for the batch systems were 80 to 85%. The transformation reaction was not first order in solution concentration or total system concentration for either of the two isomers. Measured reaction rate coefficients ( $\lambda_a$ ) and orders ( $N_a$ ) for the two compounds in experiments with initial concentrations of approximately 1850 nmol/ml were: 0.17 [nmol/hr]/[(nmol/ml) <sup>$N_a$</sup> ] ( $\ln \lambda_a = -1.79$ ) and 0.00023 [nmol/hr]/[(nmol/ml) <sup>$N_a$</sup> ] ( $\ln \lambda_a = -8.37$ ) with reaction orders 1.22 and 1.77 for *trans*-1,2-DCE and *cis*-1,2-DCE, respectively. Sorption equilibrium was apparently attained within 1.1 hr. The form of sorption could be adequately described by Freundlich-type isotherms for both compounds over the concentration range measured. The magnitude of sorption was greater for *trans*-1,2-DCE than for the more soluble *cis*-1,2-DCE. The distribution of organic products produced by the two isomers indicates some divergence in reaction pathways. While both compounds produced large proportions of ethene and ethane, transformation of *cis*-1,2-DCE resulted in significantly greater production of vinyl chloride than did *trans*-1,2-DCE.

# **REDUCTION KINETICS AND SORPTION OF *cis*- AND *trans*-1,2-DICHLOROETHENE IN A METALLIC IRON/WATER SYSTEM**

Richelle M. Allen-King and Richard M. Halket

## **INTRODUCTION**

Zero-valent iron can bring about the reduction of chlorinated solvents, such as perchloroethene and trichloroethene (PCE and TCE), dissolved in wastewater [Sweeny, 1981a; Sweeny, 1981b; Senzaki, 1991]. The reduction reaction is coupled to iron corrosion (oxidative dissolution) [Matheson and Tratnyek, 1994]. Because transformation rates are rapid relative to typical groundwater flow velocities, metallic iron is being tested in permeable *in situ* treatment barriers as a method for remediating contaminated groundwater [Gillham and O'Hannesin, 1994; Wilson, 1995].

In recent research, Burris et al. [1995] have demonstrated that sorption to nonreactive sites in iron/water systems is significant for PCE and TCE. For these two chemicals, the reduction reaction is: 1) first order in solution concentration when sorption to inert or nonreactive sites is accounted for; and 2) of a similar rate for the two compounds [Burris et al., 1995]. Low concentrations of dichloroethene (DCE) isomers and vinyl chloride (VC), as well as chloride ions, ethene, ethane, acetylene, methane, and C<sub>3</sub>-C<sub>5</sub> alkanes have been detected as reaction products [Senzaki, 1991; Gillham and O'Hannesin, 1994; Burris et al., 1995; Campbell and Burris, 1995; Burris et al., in prep].

In order to accurately plan *in situ* remediation systems, a complete understanding of the reduction reaction pathways, rates and orders are needed. The fact that some of the potential products, such as VC, are themselves toxic makes this requisite knowledge for planning application of this technology. However, the degradation rates for intermediate reaction products, such as the DCE isomers, in the iron/water system have not previously been determined. They are reduced more slowly than TCE or PCE by microorganisms [Vogel et al.,

1987].

In this research, we focus on the reduction of *cis*- and *trans*-1,2-dichloroethene (*cis*-DCE and *trans*-DCE), to contribute to the overall goal of improving our understanding of the degradation scheme for chlorinated ethenes in the metallic iron/water system. The specific objectives of this study are to determine the reduction rates, reaction order and products, and sorption of *cis*-DCE and *trans*-DCE in a metallic iron/water system. Because the 1,2-DCE isomers are frequently identified as co-contaminants with PCE and/or TCE in groundwater and have been identified independently as groundwater contaminants [Westrick, 1990], the results will also be useful in evaluating the potential for remediation of 1,2-DCEs using zero-valent iron.

## MATERIALS AND METHODS

### *Iron and chemicals*

Fisher Scientific 40-mesh sized iron filings (lot number 946315) was used for this study. All iron was pretreated in 400 g batches by washing with 1 L of 1 N HCl for one-half hour then rinsing with 10 L of argon-sparged deionized water. After rinsing, the iron was dried overnight under a nitrogen atmosphere at 90 degrees C. Once dried the iron was stored in argon sparged glass jars to inhibit oxidation. Surface area determined by BET (nitrogen) analysis was 0.9 m<sup>2</sup>/g for the untreated iron and 1.0 m<sup>2</sup>/g for the treated iron [Burris et al., 1995]. Pyrite obtained from Wards Natural Science Establishment was used to pH buffer the batch system. Pretreatment of the pyrite consisted of powdering with a mortar and pestle and storing in argon sparged glass jars to inhibit oxidation.

Both *cis*- and *trans*-1,2-DCE were supplied by Aldrich Chemical Co. with purities of 97% and 98%, respectively. Methanol was Fisher HPLC grade and hexane was Fisher Optima grade. All water used in the experiments was treated with a Milli Q reagent water system.

### ***Batch system***

Two time-series experiments were conducted with each of the two DCE isomers at two different initial concentrations in batch systems. The batch systems were prepared by adding 10.00 g of pretreated iron and 0.20 g of powdered pyrite to argon-purged 5 ml nominal (10.2 ml observed) Wheaton serum vials. DCE-amended water was dispensed to the vials in an anaerobic glove box (5% H<sub>2</sub>, balance N<sub>2</sub> atmosphere) to prevent oxidation during dispensing, and the vials were immediately sealed with teflon-lined rubber septa (Supelco). Iron- and pyrite-free controls were prepared to check for system losses attributable to mechanisms other than reduction. DCE-free controls were prepared with iron and pyrite to determine the chloride background concentration. After preparation, vials were shaken at 12 rpm at ambient laboratory temperature (22 to 25°C) until sampled. Four vials (two sample and two Fe-free control vials) were sacrificed at each sampling time for measurement of DCE solution concentrations and total DCE mass. Water samples (50 to 100 µL) were collected from the Fe-containing DCE and duplicate DCE-free control vials at each sampling time. Near the end of each time-series experiment, organic degradation products were determined by a headspace analysis, as described below.

### ***Analytical***

The aqueous DCE concentration was determined by transferring as much water as possible to a second 5 ml serum vial containing 3.00 ml hexane through a canula by air displacement. On average  $2.40 \pm 0.16$  ml of water remained in the Fe-containing vial after the transfer. The remaining DCE mass in the system, including mass sorbed, was extracted by adding 5 ml of methanol to the original batch vial and shaking for 15 minutes on a vortex mixer at high speed. Two milliliters of the resulting methanol/water mixture were diluted with 6.5 ml Milli-Q water and extracted with 4.5 ml hexane. The masses of total solution in the original vial before sampling, total methanol/water mixture, and methanol/water mixture sample transferred, were determined gravimetrically with an analytical balance (Metler). The hexane used for all extractions contained chloroform as an internal standard. Controls were treated the same way as samples with the exception that 6.5 ml of water was transferred to the receiving vial and methanol was added to the remaining water in the original sample vial. All samples were

analyzed on a Hewlett-Packard Model 5890 gas chromatograph equipped with flame ionization detector, Model 7673 autosampler, and a 60 m megabore DB-Wax (J&W Scientific) capillary column with a 1  $\mu$ m coating. The GC operating parameters were: 2.9 ml/min column flow; 11.6 ml/min split flow; 3  $\mu$ l sample injection; injector and detector temperatures were 200°C and 240°C, respectively; oven program was 50°C for 1 min, 3 °C/min to 98°C. The concentration of *cis*-DCE and *trans*-DCE were determined with a multilevel calibration curve obtained by measurement of standards with known concentration.

Near the end of each time-series experiment, a 1.00 ml headspace for analysis of volatile transformation products was created in a sample vial by displacing water with air. Following phase-equilibration at ambient temperature, a 100  $\mu$ l subsample of the headspace was analyzed using the method of Campbell and Burris [1995].

Water samples for chloride (Cl<sup>-</sup>) analysis were collected with a syringe prior to collecting the DCE water samples. Each sample was diluted with Milli-Q water to achieve a final concentration in the 0 to 20 mg/l calibration range. Samples were analyzed on a Dionex system equipped with a CDM-2 (conductivity) detector and an AS11 column. The chloride concentration was determined with a multilevel standard calibration curve.

## RESULTS AND DISCUSSION

### *Time-series and sorption*

Solution concentrations and total mass of the DCE isomers decreased throughout the course of an experiment in the Fe-containing samples (Fig. 1) while the concentrations were maintained near the initial value in the Fe-free control vials. On average 94% and 90% of the known mass added to the samples was recovered in the initial extractions for the *cis*- and *trans*-DCE isomers, respectively. These high relative recoveries (near 100%) demonstrate the success of the extraction technique in recovering the total mass added to the system. The lower recovery for the *trans*-DCE isomer is attributed to its faster reaction rate.

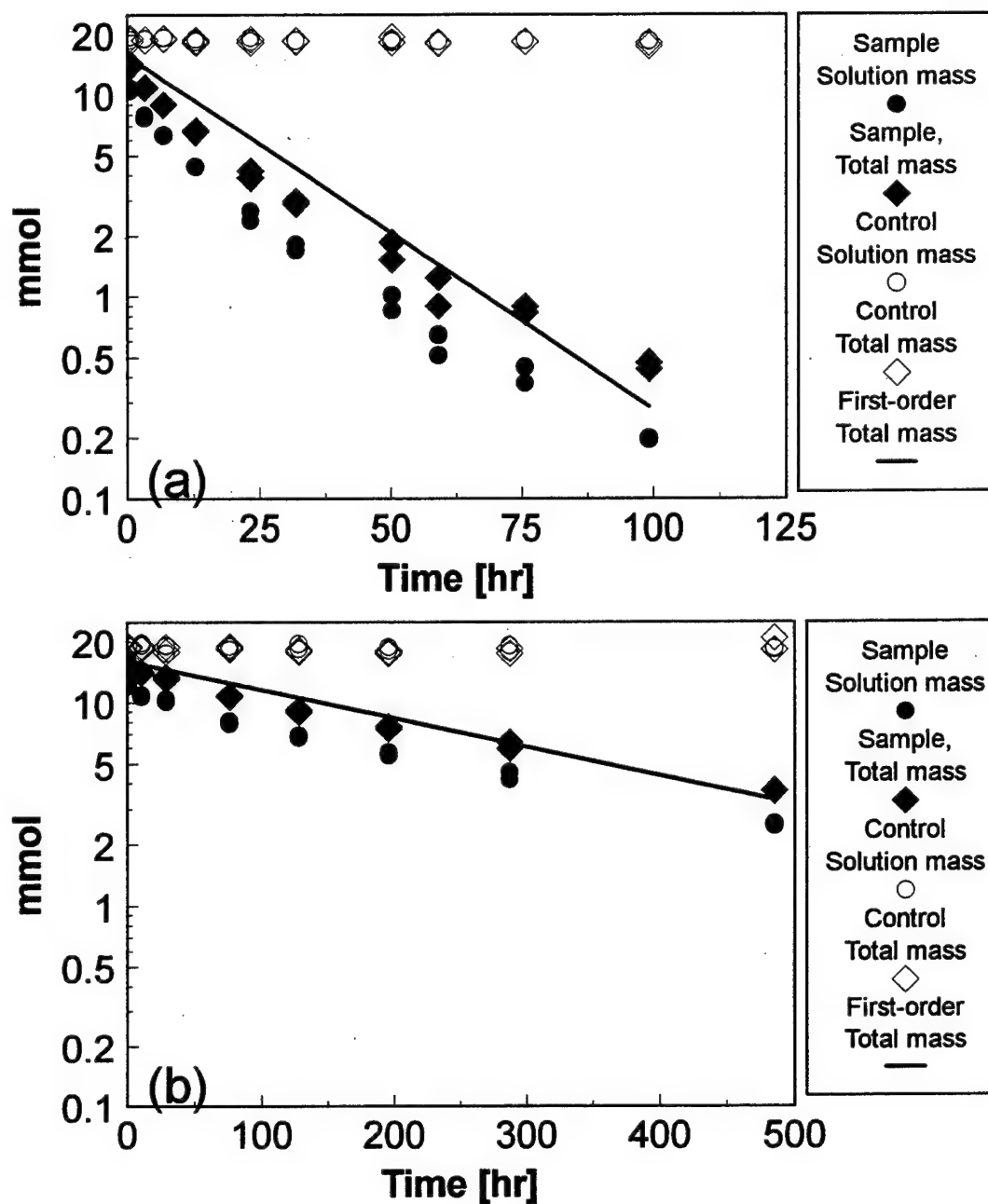


Fig. 1. Total and dissolved (a) *trans*-DCE and (b) *cis*-DCE masses over the experiment time course. *Trans*-DCE is transformed much more rapidly than *cis*-DCE. The difference between the total and dissolved masses is the sorbed mass.

Sorbed concentrations were determined for each sample from the difference between the total and solution DCE masses. Sorption isotherms for each of the two isomers are shown in Fig. 2. Sorption equilibrium was apparently achieved within 1.1 hr, as evidenced by the consistency between samples taken at 1.1 hr and at much longer times. Samples of the *trans*-DCE experiment collected after 0.5 hr were apparently not equilibrated with respect to sorption.

Sorption isotherms were fit to all equilibrium data with the Freundlich equation:

$$q_e = K_F C_a^{1/n} \quad (1)$$

where  $q_e$  is the sorbed concentration and  $C_a$  is the solution concentration. The coefficients determined are listed in Table 1. The greater magnitude of sorption for *trans*-DCE may be related to its lower solubility. The fact that the magnitude of sorption is greater for the more hydrophobic of the two solutes is consistent with previous observations. Sorption for both of the DCE isomers was much less than sorption of TCE and PCE observed in a previous study under similar conditions [Burris et al., 1995].

**Table 1.** Freundlich sorption coefficients (standard error) and number of observations ( $m$ ) for *trans*- and *cis*-DCE.

compound	$\log K_F^*$	$1/n$	$m$
<i>trans</i> -DCE	0.471(0.032)	0.685(0.011)	40
<i>cis</i> -DCE	0.469(0.029)	0.658(0.013)	34

\* Dimensions of  $K_F$  are [nmol/g]/[(nmol/ml)<sup>1/n</sup>]

#### **Transformation rate coefficients and reaction orders**

Rate coefficients ( $\lambda$ ) and reaction orders ( $N$ ) were determined for the following equations:

$$\frac{dM_T}{dt} = -\lambda_T M_T^{N_T} \quad (2)$$

and



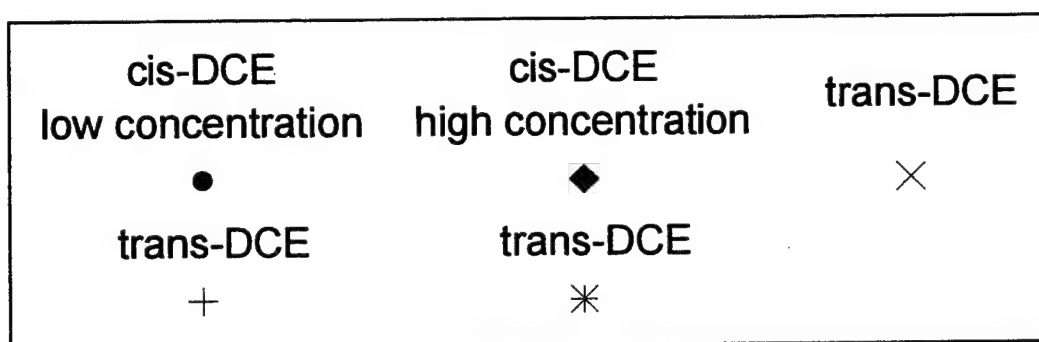
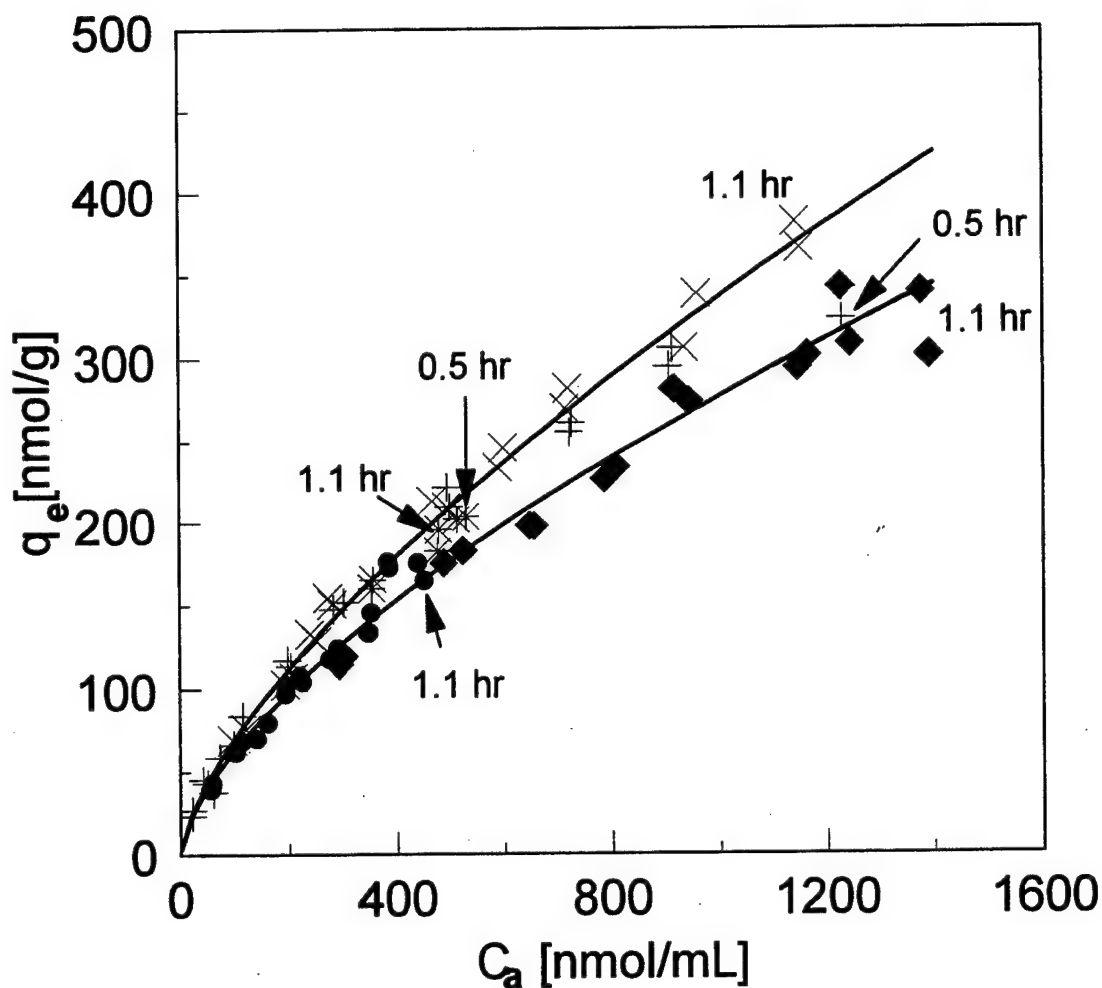


Fig. 2. Sorption isotherms. Sorbed concentration ( $q_e$ ) is determined as the difference between the sorbed and total dissolved masses at each sampling time. Sorption equilibrium occurred within 1.1 hr.

$$\frac{dM_T}{dt} = -\lambda_a C_a^{N_a} \quad (3)$$

where  $M_T$  is the total DCE mass in the vial, superscripts  $T$  and  $a$  refer to coefficients determined based on total system mass and aqueous concentration, respectively. The equations were linearized by ln-transformation and the coefficients were determined by linear regression (Fig. 3 and 4). The results of replicate samples for each time were averaged.

The rate coefficients for *cis*-DCE were much lower than those for *trans*-DCE. Little difference between the  $N_a$  and  $N_T$  for each compound, compared to the differences observed for PCE for example [Burris et al., 1995], is attributable to the relatively low sorption which was observed. A visual comparison of the fit- ( $N_T$ ) and first-order curves for *cis*-DCE is shown in Fig. 5. Clearly, the fit-order provides a better match of the measured data than the simpler first-order curve. The fit reaction orders for both DCE isomers were greater than unity for all cases, although  $N_a$  for *trans*-DCE was near unity. Previous research demonstrated that  $N_a$  was near unity for both TCE and PCE in similar experiments [Burris et al., 1995]. Therefore, our reaction order results are not consistent with the previous research, particularly for the *cis*-DCE isomer.

**Table 2.** Transformation rate coefficients ( $\lambda$ ) and reaction orders ( $N$ ) (standard error) for *trans*- and *cis*-DCE experiments.

Compound/ Experiment	$\ln \lambda_I$	$N_T$	$\ln \lambda_T$	$N_a$	$\ln \lambda_a$
<i>trans</i> -DCE	-3.22 (0.03)	1.41 (0.12)	-6.55 (0.37)	1.22 (0.11)	-1.79 (0.37)
<i>cis</i> -DCE, low	-5.57 (0.02)	1.80 (0.25)	-11.85 (0.34)	1.64 (0.24)	-6.33 (0.35)
<i>cis</i> -DCE, high	-5.74 (0.03)	1.89 (0.35)	-13.76 (0.34)	1.77 (0.33)	-8.37 (0.34)

The first-order rate coefficient ( $\lambda_I$ ) is determined from the equation (after Schwarzenbach et al., 1993)  $\lambda_I = -(\ln M_T/M_{T,0})/t$ , where  $M_{T,0}$  is the initial total DCE mass ( $t = 0$ ). Dimensions are:  $\text{hr}^{-1}$  for  $\lambda_I$ ,  $[\text{nmol/hr}]/[(\text{nmol})^{N_T}]$  for  $\lambda_T$ , and  $[\text{nmol/hr}]/[(\text{nmol/ml})^{N_a}]$  for  $\lambda_a$ .

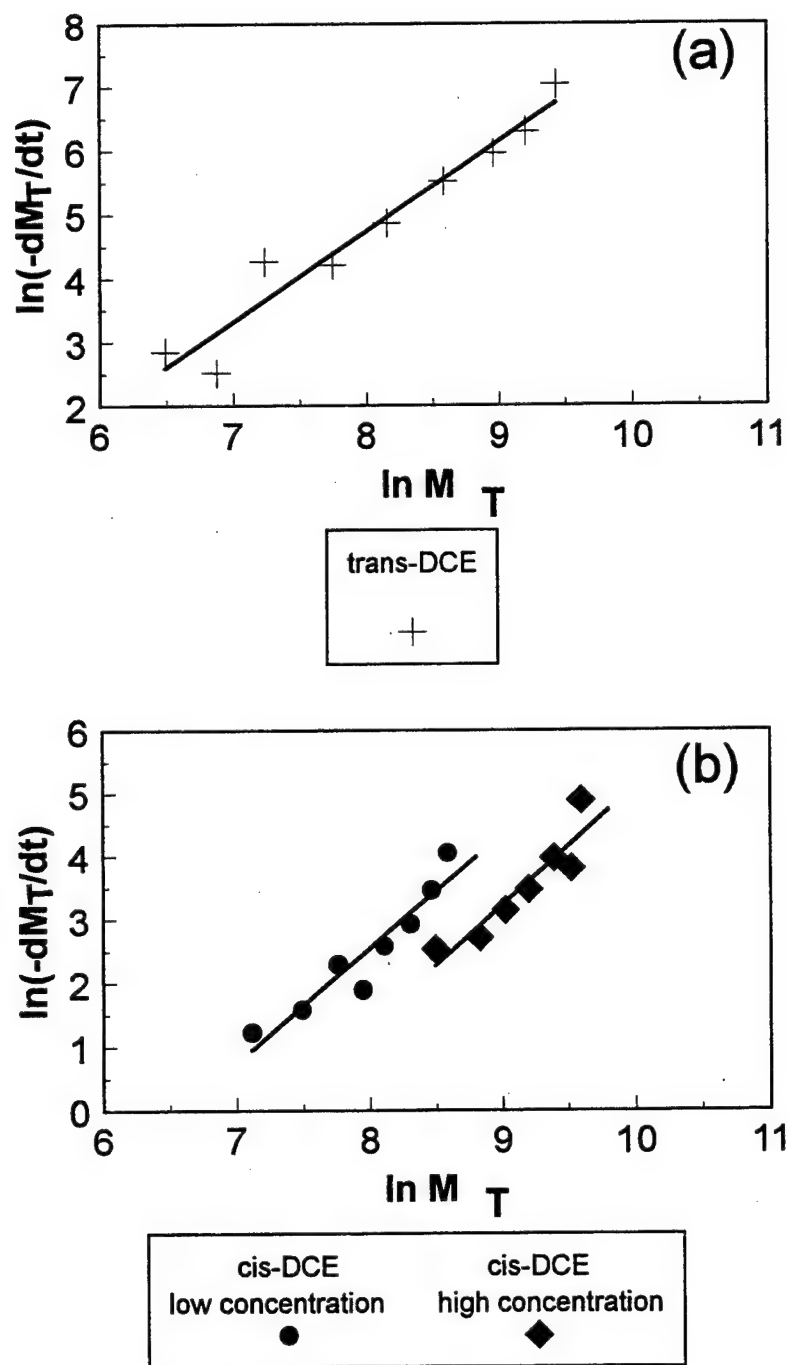


Fig. 3. Transformation rate-law determination for (a) *trans*-DCE and (b) *cis*-DCE on a total mass ( $M_T$ ) basis from linearized plots, as described in the text.

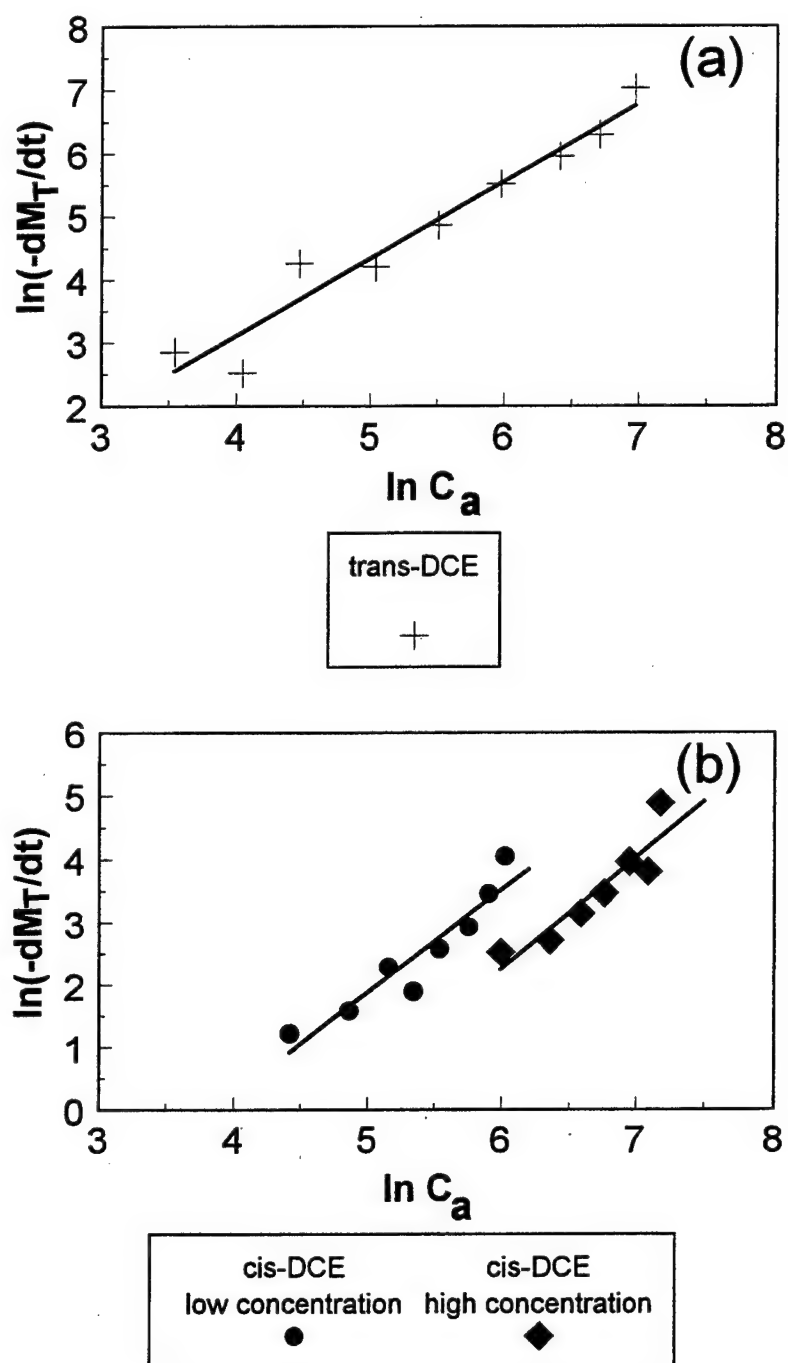


Fig. 4. Transformation rate-law determination for (a) *trans*-DCE and (b) *cis*-DCE on a solution concentration ( $C_a$ ) basis from linearized plots, as described in the text.

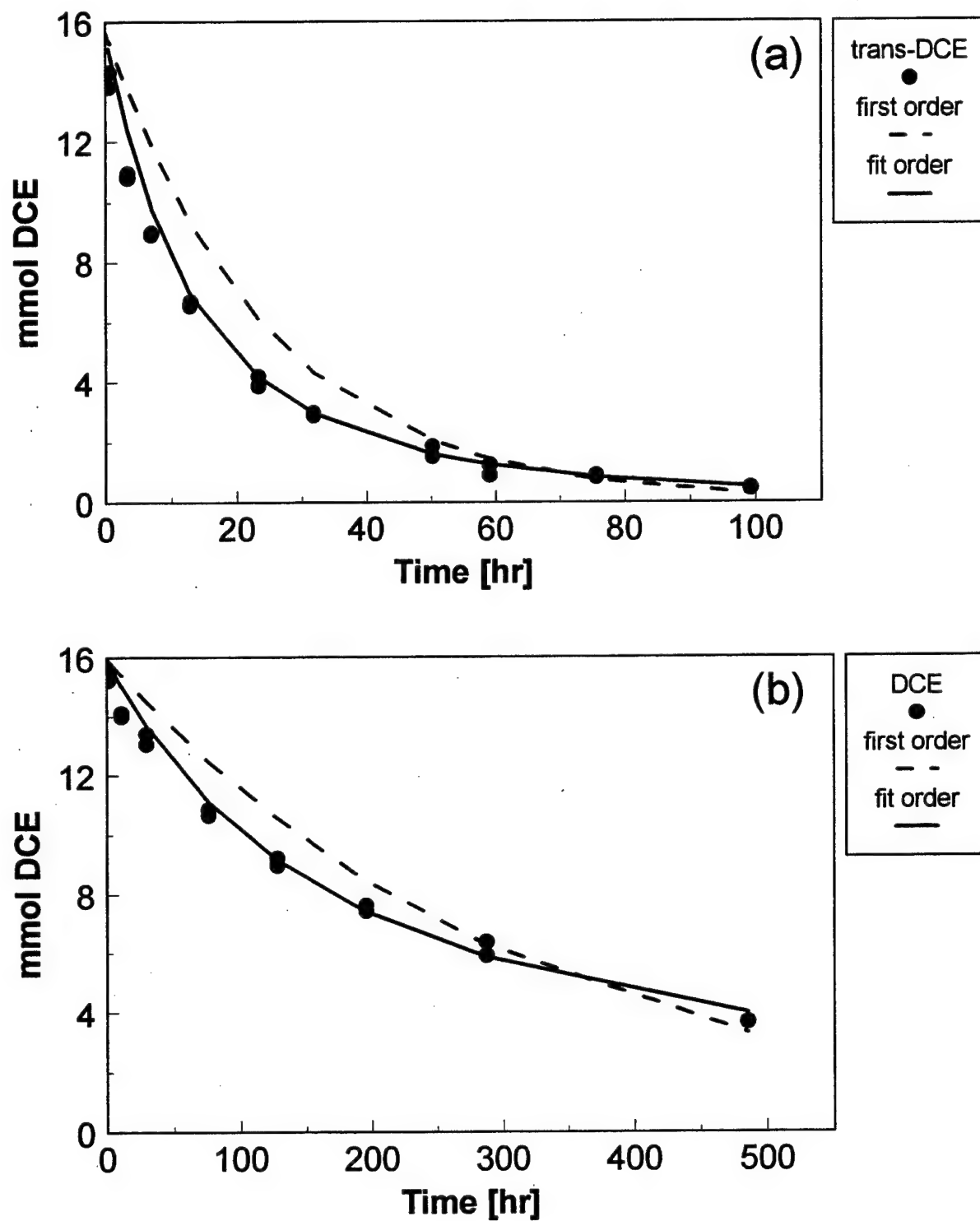


Fig. 5. Integrated first- and fit-order rates with observed total mass recovered for (a) *trans*-DCE and (b) *cis*-DCE.

### ***Reaction products and chlorine mass balance***

The organic reaction products identified near the end of the reaction course for both DCE isomers in order of decreasing predominance at the time of analysis were: ethene, ethane, C<sub>3</sub>-C<sub>5</sub> alkanes, acetylene, vinyl chloride (VC). The dominant products, by far, were ethene and ethane. Vinyl chloride is transformed slowly (Campbell, pers. comm.) relative to either of the DCE isomers in this study. As a result, VC tended to accumulate in the experiment vials as the reactions progressed. Of the DCE which had been transformed at the time of measurements, approximately 2% and <0.3% had been converted to vinyl chloride for *cis*-DCE and *trans*-DCE, respectively. The fraction converted to VC was similar for both *cis*-DCE experiments.

The chlorine balance throughout the reaction course was determined by monitoring chlorine present as either DCE or free chloride ion. Free-chloride ion increased throughout the experiments (Figs. 6 and 7). The chloride background determined in DCE-free, Fe-containing controls was about 1800 nmol (140 nmol/ml or 5 mg/L). While not unimportant in the mass balance, the background was relatively low and consistent compared to the reservoirs of chlorine atoms present in the forms of DCE or free-chloride ion. The origin of the background Cl<sup>-</sup> was either the iron metal or, perhaps more likely, the HCl used to clean the Fe. The total moles of chlorine added to each vial was estimated from the known initial DCE mass (concentration and solution volume for each vial) and the mean Cl<sup>-</sup> background. Total chlorine observed was determined as the sum of the measured moles of DCE and Cl<sup>-</sup> for each vial sampled. At the last sampling point shown, the measured chlorine accounts for 80 to 85% (mol/mol) of the total. The only chlorinated organic product identified was VC, which comprised a low proportion of the total chlorine added (<2%).

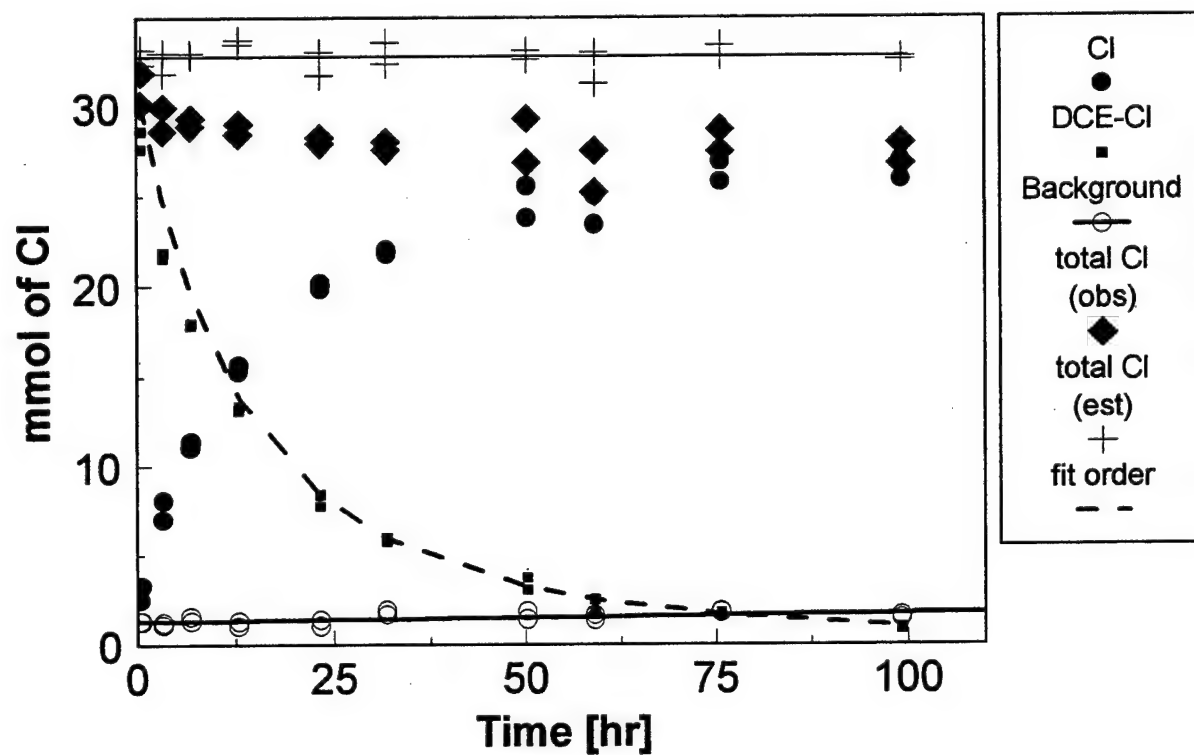


Fig. 6. Chlorine balance for *trans*-DCE. Total observed chlorine is the sum of the measured  $\text{Cl}^-$  and chlorine in untransformed *trans*-DCE. Total estimated chlorine is sum of the chlorine that was added initially as *trans*-DCE and the mean background  $\text{Cl}^-$ .

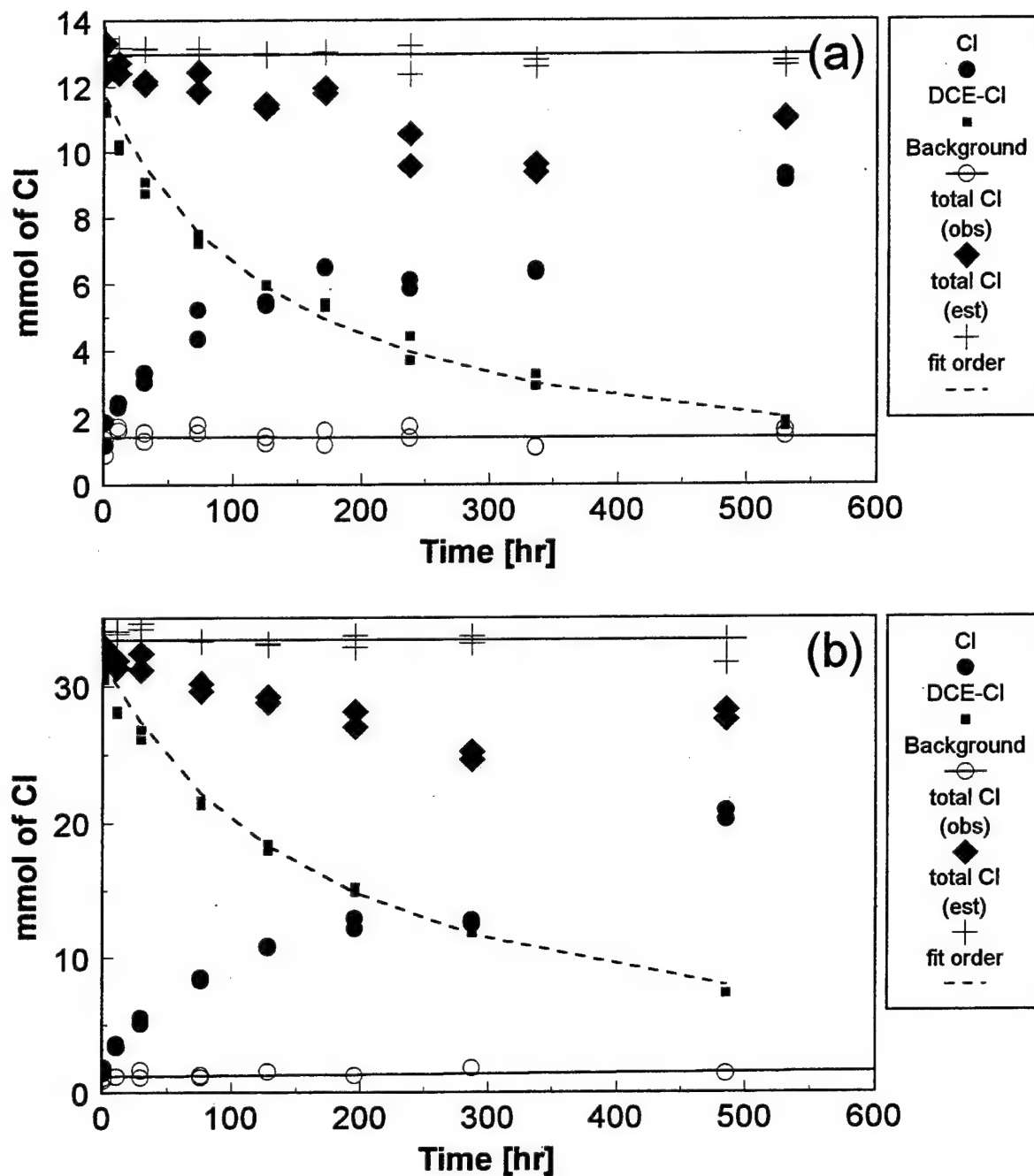


Fig. 7. Chlorine balance for (a) low concentration and (b) high concentration *cis*-DCE experiments. Total observed chlorine is the sum of the measured  $\text{Cl}^-$  and chlorine in untransformed *cis*-DCE. Total estimated chlorine is sum of the chlorine that was added initially as *cis*-DCE and the mean background  $\text{Cl}^-$ .



## SUMMARY

*Cis*-DCE and *trans*-DCE were reductively dechlorinated in time series experiments using batch systems consisting of zero-valent iron and water. Reductive dechlorination kinetics and sorption were determined for both isomers. Unlike previous results for PCE and TCE, the transformation reaction for either isomer of DCE was not first order in solution concentration or total system concentration. Measured reaction rates and orders for the two compounds in experiments with initial concentrations of approximately 1850 nmol/ml were: 0.17 [nmol/hr]/[(nmol/ml)<sup>N<sub>a</sub></sup>] ( $\ln \lambda_a = -1.79$ ) and 0.00023 [nmol/hr]/[(nmol/ml)<sup>N<sub>a</sub></sup>] ( $\ln \lambda_a = -8.37$ ) with reaction orders 1.22 and 1.77 for *trans*-DCE and *cis*-DCE, respectively. In general *trans*-DCE was degraded much more quickly than *cis*-DCE. Chloride and organic products were produced by the transformation reaction. The chlorine mass balances for the batch systems at the end of the time series experiments were between 80 to 85%. The distribution of organic products produced by the two isomers indicates some divergence in reaction pathways. While both compounds produced large proportions of ethene and ethane, transformation of *cis*-DCE resulted in significantly greater production of vinyl chloride.

The form of sorption could be adequately described by Freundlich-type isotherms for both compounds over the concentration range measured. The Freundlich sorption coefficients ( $\log K_F$ ) were 0.469 and 0.471 and the Freundlich exponents ( $1/n$ ) were 0.658 and 0.685, respectively, for *cis*-DCE and *trans*-DCE. The magnitude of sorption was greater for *trans*-DCE than the more soluble *cis*-DCE. Sorption equilibrium was apparently attained rapidly (within 1.1 hr).

## REFERENCES

1. Burris, D. R., T. J. Campbell, and V. S. Manoranjan. 1995. Sorption of trichloroethylene and tetrachloroethylene in a batch reactive metallic iron-water system. *Environ. Sci. Technol.* In press.

2. Burris, D. R., V. S. Manoranjan, T. J. Campbell, and T. M. Sivavec. Tetrachloroethylene and trichloroethylene reduction by zero-valent iron: Mass transfer limitations. In prep.
3. Campbell, T. J., and D. R. Burris. 1995. Analysis of chlorinated ethene reduction products in vapor/water phase systems by dual-column, single-detector gas chromatography. *Internat. J. Environ. Anal. Chem.* In press.
4. Gillham, R. W., and S. F. O'Hannesin. 1994. Enhanced degradation of halogenated aliphatics by zero-valent iron. *Ground Water* 32:958-67.
5. Matheson, L. J., and P. G. Tratnyek. 1994. Reductive dechlorination of chlorinated methanes by iron metal. *Environ. Sci. Technol.* 28:2045-53.
6. Schwarzenbach, R. P., P. M. Gschwend, and D. M. Imboden. 1993. *Environmental Organic Chemistry*. New York, NY: John Wiley.
7. Senzaki, T. 1991. Removal of chlorinated organic compounds from wastewater by reduction process: III. Treatment of trichloroethylene with iron powder. *Kogyo Yosui* 391:29-35.
8. Sweeny, K. H. 1981. The reductive treatment of industrial wastewaters: I. Process description. In *Symposium Series 209*, ed. G. F. Bennett, 67-71. Vol 77. American Institute of Chemical Engineers.
9. Sweeny, K. H. 1981. The reductive treatment of industrial wastewaters: II Process applications. In *Symposium Series 209*, ed. G. F. Bennett, 72-78. Vol 77. American Institute of Chemical Engineers.
10. Vogel, T. M., C. S. Criddle, and P. L. McCarty. 1987. Transformations of halogenated aliphatic compounds. *Environ. Sci. Technol.* 21:722-36.

11. Westrick, J. J. 1990. National surveys of volatile organic compounds in ground and surface waters. **In** *Significance and Treatment of Volatile Organic Compounds in Water Supplies* N. M. Ram (ed.), 103-25. Chelsea, MI: Lewis.
12. Wilson, E. K. 1995. Zero-valent metals provide possible solution to groundwater problems. *Chemical & Engineering News*, July 3, 19-22.

**EPIDEMIOLOGICAL IDENTIFICATION OF A METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS* OUTBREAK IN KENTUCKY EMPLOYING REP-PCR IN COMPARISON TO PFGE**

**Catherine J. Healy  
Graduate Research Associate  
Department of Molecular Biology**

**University of Scranton  
Loyola Hall, Wyoming Avenue  
Scranton, PA 18510**

**Final Report for:  
Summer Graduate Research Program  
Armstrong Laboratory**

**Sponsored by:  
Air Force Office of Scientific Research  
Brooks Air Force Base, TX**

**and**

**Armstrong Laboratory**

**August 1995**

EPIDEMIOLOGICAL IDENTIFICATION OF A METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS* OUTBREAK IN KENTUCKY EMPLOYING REP-PCR IN COMPARISON TO PFGE

Catherine J. Healy  
Graduate Research Associate  
Department of Molecular Biology  
University of Scranton

Abstract

*Staphylococcus aureus* is a major cause of both nosocomial and community-acquired infections. An ongoing surveillance system which would rapidly and accurately identify clonal relationships between outbreaks would better equip physicians and epidemiologists in treating patients with the appropriate antibiotic regimen. This study categorizes and identifies 42 clinical isolates of methicillin-resistant *Staphylococcus aureus* from various body sites such as sputum, open wounds, and joints. Results from repetitive sequence-based polymerase chain reaction yield three distinct strains.

# EPIDEMIOLOGICAL IDENTIFICATION OF A METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS* OUTBREAK IN KENTUCKY EMPLOYING REP-PCR IN COMPARISON TO PFGE

Catherine J. Healy

## Introduction

The prevalence of nosocomial and community-acquired infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA) is increasing at an alarming rate. The implication of a global surveillance system specifically to combat antibiotic resistance has been suggested (5). Numerous epidemiologic markers have been useful for identifying sources and monitoring the spread of these infections. Phenotypic systems such as antibiotype, biotype, phage type, ribotyping, plasmid profile, immunoblotting, structural protein or enzyme electropherotype have limitations in discriminatory ability (8). Pulsed-field gel electrophoresis (PFGE) and repetitive sequence-based polymerase chain reaction (rep-PCR) analyze subtle changes in genetic variation between organisms, and are therefore more sensitive methods of typing. PFGE has been recommended as the typing method of choice for MRSA by several groups of investigators (3,6,7). Other investigators acknowledge that although PFGE is more discriminatory, rep-PCR is a more rapid, cost effective, and reproducible procedure that yields sufficient discrimination between strains (9).

PFGE is the process in which organisms are embedded in agarose, lysed *in situ*, and digested with restriction endonucleases that cleave infrequently. Various fragments are resolved into discrete bands using an electrophoresis method called contour-clamped homogenous electric field (CHEF) mapping. At the Third International Meeting on Bacterial Epidemiological Markers, April 6-9, Cambridge, U.K., guidelines for PFGE interpretation have been established. These guidelines can be seen in Table 1.

Prokaryotic genomes are known to contain repetitive palindromic sequences interspersed between regions of non-repetitive sequences (9). These repetitive sequences are conserved or common among many gram positive bacteria, whereas the non-repetitive sequences vary among bacteria of different species and strain. Rep-PCR exploits these short polynucleotide sequence patterns.

With the selection of oligonucleotide primers, complementary to the interspersed repetitive sequences, various sized DNA fragments of the non-repetitive sequences will be amplified in a PCR reaction. When amplified DNA fragments are resolved by agarose gel electrophoresis, distinguishable fingerprints are produced specifically for each strain. Rep-PCR is a more rapid, cost effective and reproducible procedure than PFGE.

### Materials and Methods

**Culturing.** Clinical isolates of methicillin-resistant *Staphylococcus aureus* infections from sputum, open wounds, urine and stool were obtained from Dr. Martin Evans of the Department of Internal Medicine at the University of Kentucky, Lexington, Kentucky. Cultures were streaked on 5% blood agar plates (Remel, Lenexa, KS) and incubated over night in a 37°C incubator. Single colony isolates were selected and grown in 3 ml of brain heart infusion (BHI) broth over night at 37°C with agitation.

**Isolation.** The method of isolation is modeled from Ausubel (1), with modifications. Over night cultures were pelleted at 3,000 x g for 10 min. Pellets were resuspended in 500- $\mu$ l of 10 mM Tris-HCL-1mM EDTA (pH 8.0), and transferred to a 1.5 ml eppendorfs tube. The cells were lysed by adding 5- $\mu$ l of 4 mg/ml lysostaphin followed by a one hour incubation at 37°C. In order to digest excess RNA, 5- $\mu$ l of 10 mg/ml of RNase A (Sigma Chemical Company, St. Louis, MO) was added to the solution followed by a 15 min incubation at 37°C. In order to digest excess protein and cell debris, 10- $\mu$ l of 20 mg/ml proteinase K (Ameresco, Solon, OH) and 150- $\mu$ l of 10% SDS were added to the solution and incubated for another hour at 37°C. An equal volume of phenol:chloroform (Ameresco) was added to the solution, mixed by inversion and centrifuged at 14,000 x g for 10 min. The top, aqueous layer was drawn off using a pasteur pipet and transferred to a clean 1.5 ml eppendorf tube. An equal volume of chloroform (Ameresco) was added to this solution, mixed by inversion, and centrifuged at 14,000 x g for 10 min. Again the top, aqueous layer was drawn off and transferred to a clean 1.5 ml eppendorf tube. DNA was precipitated from this solution by adding an equal volume of 100% isopropanol, and centrifuged at

14,000 x g for 5 min. The isopropanol was poured off, and DNA pellets were allowed to air dry. DNA was resuspended in 100- $\mu$ l of sterile distilled water.

**Quantitation.** DNA isolations were quantitated by visualization on a 1% Sea Kem GTG agarose (FMC Bioproducts, Rockland, Maine) in 89 mM Tris-borate-2 mM EDTA (pH 8.0) containing ethidium bromide (0.5  $\mu$ g/ml). Approximately 200 ng of DNA template is required in the PCR reaction. Conveniently, 200 ng of DNA is approximately the minimum amount of DNA that can be visualized on an agarose gel. Quantitations were estimated accordingly.

**PCR.** The primer used in this PCR reaction, RW3A (5'-TCGCTCAAACAACGACACC-3'), was derived from the *M. pneumoniae* repetitive sequence, RepMP3, and revised by DeVecchio (4). The 50- $\mu$ l master mix for the reaction incorporated approximately 200 ng of DNA template, 75 pmol of primer, 10 mM dNTP mix (Boehringer Mannheim, Indianapolis, MN), 2.5 U amplitaq polymerase (Perkin-Elmer Cetus, Norwalk, Conn.), 8.0- $\mu$ l of 25mM MgCl<sub>2</sub> (Perkin-Elmer), and 5- $\mu$ l of 10X PCR II buffer (Perkin-Elmer). The thermal profile for the GeneAmp PCR System 9600 thermocycler (Perkin-Elmer) included a predenaturation at 95°C for 3 min, followed by 30 cycles of 94°C for 1 min, 54°C for 1 min, and 72°C for 2 min, followed by a final extension at 72°C for 5 min.

**Electrophoresis.** Fingerprints of each sample were distinguished by loading 15- $\mu$ l of amplicon and 3- $\mu$ l of loading buffer onto a 1.8% SeaKem GTG agarose (FMC bioproducts) in 89 mM Tris-borate-2 mM EDTA (pH 8.0). A one hundred base pair ladder (Gibco BRL Life Technologies, Inc., Bethesda, MD) was run in the first lane of every gel in order that the size of the DNA fingerprint banding patterns be determined. The gel was electrophoresed for approximately 2-1/2 hours at 65 V.

**PFGE.** Dr. Martin Evans from the Department of Internal Medicine at the University of Kentucky, Lexington, Kentucky will perform the pulsed-field gel electrophoresis in this comparison study. The GenePath Group 1 Reagent Kit (BioRad, Foster City, CA), followed by restriction endonuclease digestion with *Sma* I will be used in order to generate the MRSA fingerprints. Results will be compared to rep-PCR results when they are obtained.



## Results

Cultures grew well in BHI broth as expected. The isolation technique yielded high quantities of DNA, as seen on a 1% SeaKem GTG agarose gel of the genomic DNA isolation (Fig. 1). Bright smears appear in each well containing only 3- $\mu$ l of the genomic isolation and 2- $\mu$ l of loading buffer. The DNA was quantitated according to its appearance on gels such as this. To 30- $\mu$ l of the PCR master mix, a 20- $\mu$ l dilution of 200 ng of genomic template DNA was added. The amplicon seen in Figures 2-6 were generated by PCR as described earlier. Each gel in Figures 2-6 is a 1.8% SeaKem GTG agarose gel loaded with a one hundred base pair molecular weight marker in the first lane, followed by the MRSA fingerprints (15- $\mu$ l PCR amplicon and 3- $\mu$ l of loading buffer). Eight fingerprints of MRSA type 1 are seen in Figure 2. Six fingerprints of MRSA type 1, two fingerprints of MRSA type 2, and two fingerprints of type 3 are seen in Figure 3. The remainder of the MRSA type 3 fingerprints are seen in Figures 4-6. An overall summary of the three distinguishable strains present in this outbreak as well as the body site the samples originated from is seen in Table 2.

Type 1 differs from Type 2 by  $\geq 3$  genetic events as can be seen in Fig. 3. There is a very distinguished band in Type 2 at approximately 950 bp. Type 2 also lack a very distinguished band present in Type 1 at approximately 1850 bp. Type 3 is similar to Type 1 however Type 3 lacks a distinguished band at approximately 550 bp.

## Discussion

The fingerprints seen in Figures 2-6 are indicative of the amplified variable regions specific to each strain. PCR amplicons generated by the use of primer RW3A, yielded highly discriminatory fingerprints. Three banding patterns were observed in this study. Three strains of methicillin-resistant

*Staphylococcus aureus* are believed to have caused this particular outbreak. Upon optimization of this technique, this study has proven that rep-PCR is a rapid, cost effective and reproducible procedure that is applicable in monitoring the spread of infectious MRSA. With this information available, physicians and epidemiologists can identify highly infectious strains, and treat them with the appropriate antibiotic regime before they become a widespread outbreak.

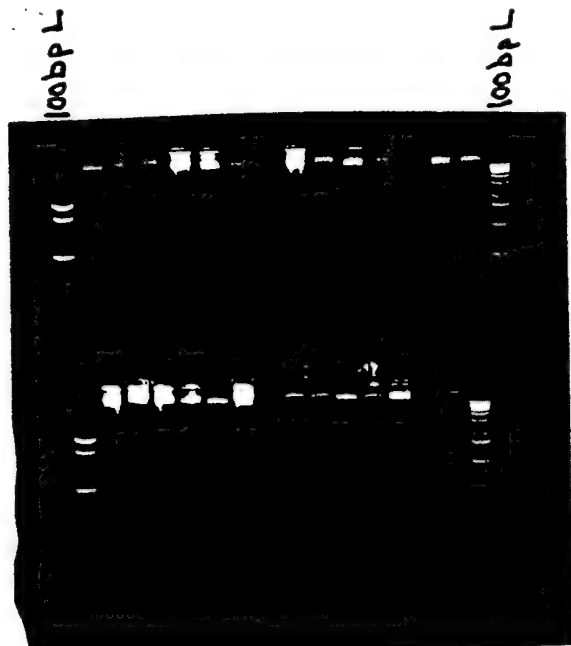
Although final results on the number of Types obtained when analyzing this outbreak using PFGE have not yet been reported, it will be interesting to compare the discriminatory ability of rep-PCR in contrast to PFGE.

Category	No. Genetic Differences compared With Epidemic Pattern	No. Band Differences Compared With Epidemic Pattern	Interpretation
Identical	0	0	Is part of the epidemic
Closely Related	1	2-4	Probably part of the epidemic
Not Closely Related	2	4-8	Much less likely to be part of the epidemic
Different	$\geq 3$	$> 8$	Not part of the epidemic

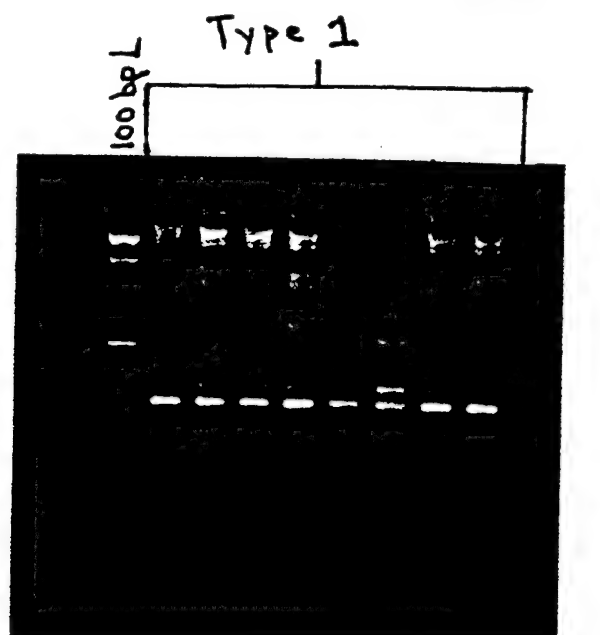
**Table 1:** Guidelines for Interpretation of Chromosomal Patterns by Pulsed-Field Gel Electrophoresis. (Unpublished article. Goering, R.V., Creighton University School of Medicine, Omaha, Nebraska).

Body Site	Type 1	Type 2	Type 3
urine	4	-	2
stool	-	-	1
wound swab	6	2	10
sputum	4	-	12

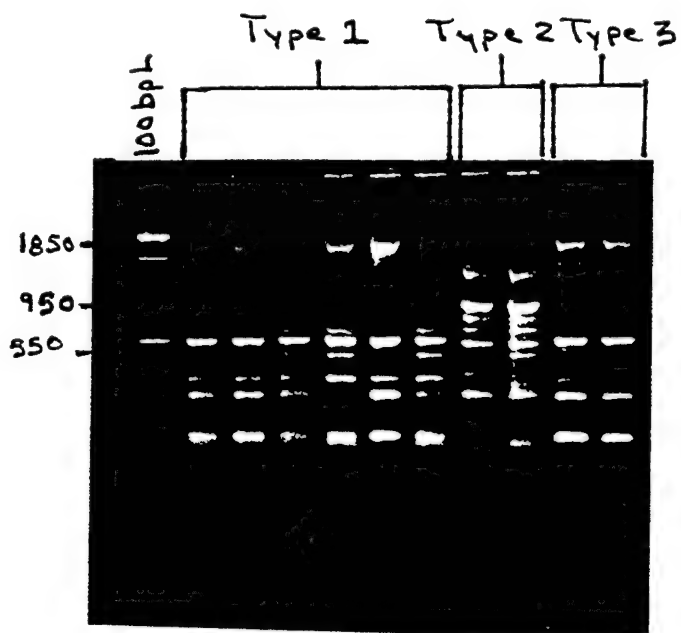
**Table 2:** Summary of Three types or strains of MRSA in the outbreak and from what body site the samples originated from.



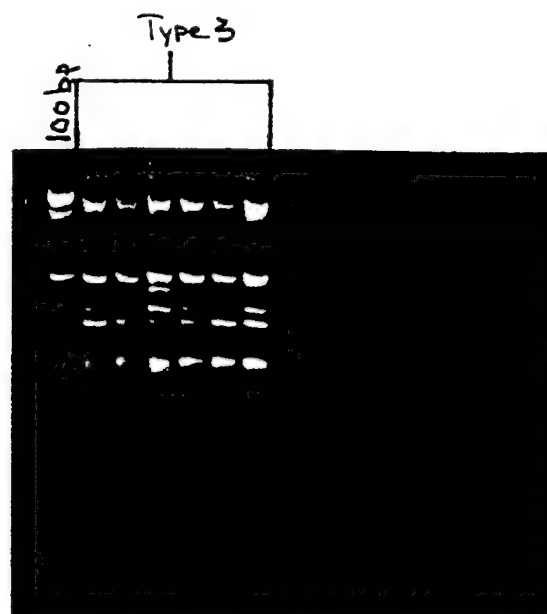
**Figure 1:** MRSA Genomic Isolation  
1% Agarose Gel (3- $\mu$ l isolation  
amplicon)



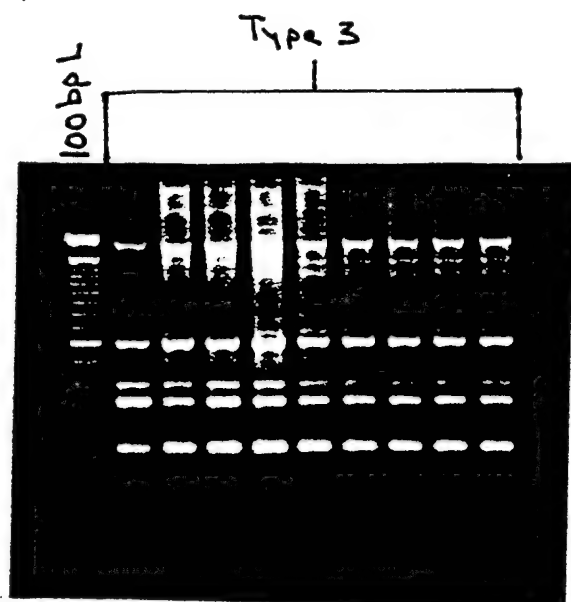
**Figure 2:** MRSA Fingerprints Type 1  
1.8% Agarose Gel (15- $\mu$ l PCR



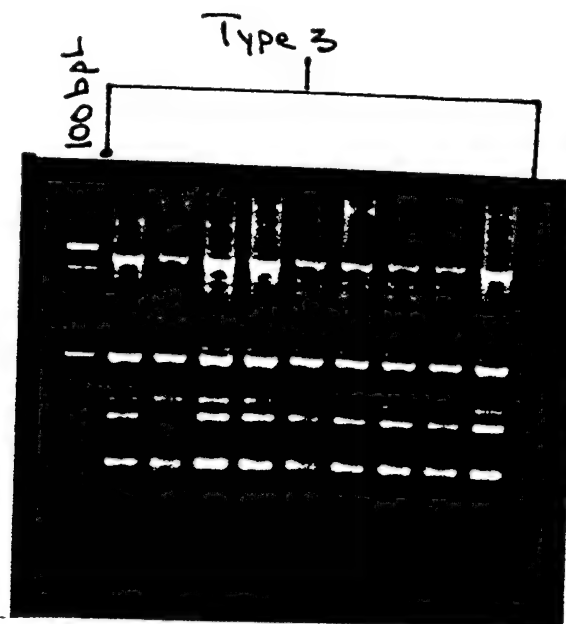
**Figure 3:** MRSA Fingerprints Types 1, 2 & 3  
Agarose Gel (15- $\mu$ l PCR amplicon)



**Figure 4:** MRSA Fingerprints Type 1.8%  
1.8% Agarose Gel (15- $\mu$ l PCR amplicon)



**Figure 5:** MRSA Fingerprints Type 3  
1.8% Agarose Gel (15- $\mu$ l PCR amplicon)



**Figure 6:** MRSA Fingerprints Type 3  
1.8% Agarose Gel (15- $\mu$ l PCR amplicon)

## REFERENCES

- (1) Ausubel, F. et al. (1991). Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley Interscience, New York.
- (2) Bustamante, C. et al. (1993). Towards a Molecular Description of Pulsed-Field Gel Electrophoresis. Trends in Biotechnology. 11: 23-30.
- (3) Carles-Nurit, M.J. et al. (1992). DNA Polymorphisms in Methicillin-Susceptible and Methicillin-Resistant strains of *Staphylococcus aureus*. J. Clin. Microbiol. 31: 982-985.
- (4) DelVecchio, V.G. et al. (1995). Molecular Genotyping of Methicillin-Resistant *Staphylococcus aureus* via Fluorophore-Enhanced Repetitive-Sequence PCR. J. Clin. Microbiol. 33: 2141-2144.
- (5) Report of the ASM Task Force on Antibiotic Resistance. (1995). American Society for Microbiology. Wash. DC.
- (6) Saulnier, P. et al. (1993). Random Amplified Polymorphic DNA Assay is Less Discriminant than Pulsed-Field Gel Electrophoresis for Typing Strains of Methicillin-Resistant *Staphylococcus aureus*. J. Clin. Microbiol. 31: 982-985.
- (7) Schlichting, C. et al. (1993). Typing of *Staphylococcus aureus* by Pulsed-Field Gel Electrophoresis, Zymotyping, Capsular Typing, and Phage Typing: Resolution of Clonal Relationships. J. Clin. Microbiol. 31: 227-232.
- (8) Struelens, M.J. et al. (1992). Epidemiologic Typing and Delineation of Genetic Relatedness of Methicillin-Resistant *Staphylococcus aureus* by Macrorestriction Analysis of Genomic DNA by Using Pulsed-Field Gel Electrophoresis. J. Clin. Microbiol. 30: 2599-2605.
- (9) Versalovic, J. Et al. (1994). Genomic Fingerprinting of Bacteria Using Repetitive Sequence-Based Polymerase Chain Reaction. Methods in Molecular and Cellular Biology. 5: 25-40.

COMPARATIVE EFFECTS OF DYNAMIC AND STATIC STRENGTH TRAINING ON +G<sub>z</sub>  
TOLERANCE

Christopher M. Hearon  
Instructor  
Department of Kinesiology

Louisiana State University  
121 Long Field House  
Baton Rouge, LA 70803-7101

Final Report for:  
Summer Faculty Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

July 1995

# COMPARATIVE EFFECTS OF DYNAMIC AND STATIC STRENGTH TRAINING ON +G<sub>Z</sub> TOLERANCE

Christopher M. Hearon  
Instructor  
Department of Kinesiology  
Louisiana State University

## Abstract

The comparative effectiveness of generalized dynamic resistance strength training and specific static resistance strength training in enhancing relaxed, gradual onset rate (GOR) +G<sub>Z</sub> tolerance and +4.5 to + 7.0 G<sub>Z</sub> simulated air combat maneuver endurance is the focus of the present investigation. Because of the muscular efforts demanded of individuals during repeated performance of the AGSM during high +G<sub>Z</sub> conditions, physically untrained individuals will fatigue earlier than their physically trained counterparts. Thus, the protective effect of general strength training programs on +G<sub>Z</sub> endurance has indicated enhancement of +G<sub>Z</sub> endurance in strength-trained subjects. Additionally, results by the Russians indicated that static force generated by a pilot during a progressive test of sustained, static leg press was highly predictive of +G<sub>Z</sub> tolerance. Thus, it was the purpose of the present investigation to compare these training regimes as to serve as an important step in responding to the flying community's request for a physical training program which enhances +G<sub>Z</sub> endurance and also reduces aircrew training time involvement. The present investigation has not been completed, and is ongoing, thus no final results are reported at present.



# COMPARATIVE EFFECTS OF DYNAMIC AND STATIC STRENGTH TRAINING ON +G<sub>z</sub> TOLERANCE

Christopher M. Hearon

## Introduction

G-tolerance is the ability to maintain consciousness, and effective performance when under G-stress. In doing so, blood pressure and blood flow must be maintained. Factors which may effect the individual's G tolerance include: 1) the effectiveness of the anti-G-straining maneuver, 2) physical, physiological and psychological factors, 3) equipment and, 4) G- discipline and knowledge. (AFPAM 11-404, p.4).

The anti-G straining maneuver (AGSM) has two principal components (Gillingham 1988). The first component is a muscular component, which requires the sustained, static recruitment of the skeletal musculature. Second, the intra-thoracic pressure component of the AGSM involves repeated 3-s cycles of forceful expiratory efforts against a closed glottis (L-1 maneuver, Cote et al. 1986, Gillingham 1988). The L-1 AGSM is the best G-defense available for aircrew members. Equipment was never meant to replace the AGSM, only aid it. Therefore, the repetition of an effective AGSM is critical to fighter aircrew in the sustained high +G<sub>z</sub> environment of air combat maneuvering.

Physical fitness programs in general, have been shown to enhance the individual's G-tolerance (Bulbulian et al. 1994, Burton 1986, 1987, Tesch et al. 1983) perhaps by enhancing the active muscles involved in the AGSM. More specifically, research has demonstrated weight training (i.e., anaerobic training) increases the individual's ability to withstand high G-forces for a longer duration with less fatigue. In addition, aerobic training has been shown to decrease the recovery time between centrifuge training runs. The present investigation appears to be critical to better understand the value of a specific strength training program on the individual's ability to withstand G-stress.

## Methodology

Equipment and Facilities. Subjects' relaxed  $+G_z$  tolerance and +4.5 to +7.0  $G_z$  SACM endurance will be assessed from standards measurement equipment during exposures on the Armstrong Laboratory Centrifuge. Subjects will be experienced members of the acceleration subject panel and will be fitted with the standard CSU-13 B/P anti-G suit, HGU-55/P helmet and MBU-12/P oxygen mask for all centrifuge runs. To minimize breathing resistance from the oxygen mask during the AGSM, a high-flow CRU-93 oxygen regulator, in the non-pressure breathing setting, will be used for all centrifuge runs.

All subjects will be fitted with a 5-lead ECG for two-axis recording of heart rhythms and heart rate during all centrifuge exposures. Analog data signals of physiologic and anti-G suit pressure variables will be recorded on a multiple channel strip recorder (Gould 2800S), with magnetic tape back-up, for subsequent data reduction and statistical analysis. Centrifuge subjects will be observed via closed circuit video monitors in the centrifuge control room.

An ACES II upright seat ( $13^\circ$  seatback angle), such as used in the F-15 aircraft, has been modified for use as the static strength training device. This "static trainer" is equipped with individual foot boards fitted with strain gauge transducers which interface with a microprocessor, allowing continuous measurements of static leg press efforts (displayed in kg) during static strength testing and strength training sessions.

With the exception of biceps curl exercises performed with barbells, subjects assigned to the dynamic, strength training program will be trained and tested on a multi-station exercise apparatus (Universal Gym). A noninvasive arterial blood pressure monitoring system (Finapres) also will be used for all subjects during tests of static and dynamic leg press strength.

A cycle ergometer (Monark 818E) will be used for subject performance of tests of aerobic capacity ( $VO_2$  peak). Peak anaerobic power and mean anerobic power will be determined from administration of a 30-s supramaximal pedaling test on a different cycle ergometer (Cardionics, Sweden).  $VO_2$  peak and respiratory exchange ratio (RER) values will be determined from analysis

of expired gas (Sensormedics 2900z Metabolic cart) collected from subjects during tests of anaerobic capacity and anaerobic power. Cycle ergometry tests will be conducted in the Exercise Physiology Laboratory in Building 110.

A pressure manometer (Boehringer) with negative inspiratory force monitor and pressure gauge adapter (Instrumentation Industries, Inc.) will be used to measure maximal expiratory pressure and maximal (negative) inspiratory pressures of all subjects. These tests will be performed by pulmonary function technicians in Building 110.

**Subjects.** Male and female subjects from the Armstrong Laboratory acceleration subject panel who have maintained acceleration training standards will be recruited for this study. We will recruit sufficient subjects to have approximately 14 subjects complete the study. Subjects must have recent +G<sub>z</sub> exposure experience and be able to complete at least four cycles (135-s) of the +4.5 to +7.0 G<sub>z</sub> SACM to qualify for the study. Because acceleration panel subjects have not routinely performed SACM exposures with helmet and mask, subjects in the study will perform SACM familiarization runs wearing the full equipment ensemble, including helmet and mask, prior to establishment of maximal baseline SACM endurance times. Panel subjects will also receive exercise program indoctrination before they perform the strength testing and initiate the respective 8-week training regimens. Only acceleration panel subjects who have no history of prolonged strength training participation and have not participated in muscular strength training programs during the past four weeks will be studied. To control for the potentially detrimental effects of excessive aerobic training on +G<sub>z</sub> tolerance (Bulbilian et al. 1994), subjects who perform regular aerobic activities equal or equivalent to running more than 15 miles per week also will not be included in the study.

**Duration of the study.** Determination of relaxed +G<sub>z</sub> tolerance, maximal SACM endurance, static or dynamic strength (with Finapres blood pressures), aerobic capacity, anaerobic power, and maximal inspiratory and expiratory pressures will require approximately one week. These determinations will be performed prior to (pre) following (post) each 8-week strength training regimen. Subjects will participate in both strength training regimen (cross-over

experimental design), with a 4-week detraining period between regimens, beginning after completion of post training testing for the first 8-week period. Duration of participation for each individual subject will be approximately 22 weeks.

General design and Overview. The general experimental design is a 2 x 2 repeated measures factorial design, blocked on subjects. There are two levels of treatment (general/dynamic & specific/static) and two levels of tests (pre & post) for measurement of physiologic variables accomplished prior to and following each 8-week training regimen. Analysis of variance (ANOVA) and tests of pairwise comparisons will be used to test for significant change in mean scores of dependent variables over the 8-week training regimens. Dependent variables include:

- a. relaxed gradual (0.1 G/s) onset rate (GOR) +G<sub>Z</sub> tolerance;
- b. peak heart rate response to +4.5 to +7.0 G<sub>Z</sub> SACM (6 G/s)
- c. SACM endurance (s)
- d. subjects reported perceived effort (scale 1-10) during SACM endurance runs
- e. aerobic capacity (VO<sub>2</sub> peak)
- f. peak anaerobic power and mean anaerobic power.
- g. static muscular strength from the static leg press test
- h. dynamic muscular strength (1 RM; in kg) and;
- i. systolic and diastolic blood pressures obtained with Finapres during both static and dynamic tests of leg press strength.

In order to maintain their entry AGSM proficiency and +G<sub>Z</sub> endurance baseline, subjects will be required to perform a +4.5 to +7.0 SACM centrifuge training profile at least once but not more than twice weekly. This +G<sub>Z</sub> training requirement consists of a rapid onset rate (ROR) 5 G/10-s warm-up, followed by four cycles (@135-s) of the SACM. Pre-and post-regimen tests of maximal SACM endurance will satisfy the weekly SACM proficiency requirements for those particular weeks. Weekly AGSM proficiency centrifuge exposures will continue during the respiratory muscular training regimen.

Day 1 of testing/subject orientation involves establishment of relaxed +G<sub>z</sub> tolerance and +4.5 to +7.0 G<sub>z</sub> SACM endurance on the centrifuge. On day 2, subjects will be tested for aerobic capacity (VO<sub>2</sub>) peak on a cycle ergometer. Following an orientation to the strength training programs on day 3 subject will be tested on day 4, for anaerobic power using an electrically braked cycle ergometer (Bar-Or 1978) . Finally, on the morning of day 5, maximal expiratory and inspiratory pressures will be tested in the pulmonary function facilities of Building 110. These tests will establish subject inspiratory and expiratory pressure baselines and be used to indirectly assess changes in respiratory muscle strength after completion of the 8-week strength training periods. Following day 5 pulmonary testing, subjects will perform a final rehearsal of the strength training regimen. At least two but not more than seven days following the various testing and strength training orientation sessions, subjects will initiate either of the two 8-week muscular strength training programs. Subjects tests and orientation will be performed in the order described.

Subjects next will be assigned in a balanced, randomized order, to either the dynamic strength training regimens or the static strength training regimen. An exercise supervisor will indoctrinate all subjects on technique training for their particular exercise regimen prior to testing subject performance on regimen-specific strength tests. With the exception of the use of dunbbells for testing (and training) of biceps strength and a 1-minute "abdominal crunch" test to assess abdominal strength, dynamic strength testing will consist of determining each subject's one repetition maximum (1RM) lifting capability for each lift on the strength training apparatus. In addition to the abdominal and neck exercises previously described, subjects in the dynamic exercise program will perform a typical body weight-training exercise on a Universal Gym apparatus. Specific exercises are as follows:

leg press

knee extension

knee flexion (leg curl)

supine bench press

arm (triceps) push-down

lat (latissimus) pull-down

Subjects static strength will be determined from the best of three maximal static leg press efforts (5-s duration) performed in the static trainer. As the second measure of static strength (static endurance), subjects will be tested on the duration (0-150 s) they can sustain the progressive 5-level static strength training task in the static trainer.

Dynamic and static strength sessions will be conducted three days per week, on alternate days. Dynamic strength training will consist of each subject's performing to fatigue on weight loads established during the first days indoctrination session. Weightloads for each exercise will be calculated as 80% of the individual's 1RM, a weight which theoretically can be lifted no more than 5-7 repetitions (5-7RM). For each static strength training session, subjects will perform a single 150-s "set" (5 x 30-s phases) of a sustained, progressively more difficult static leg press in the static training device interfaced to a computer display.

Within each 8-week training regimen, if a subject fails to participate in all three of the supervised exercise sessions per week on more than one occasion, or if the individual fails to participate in less than two sessions per week on any occasion the subject will be disenrolled from the study. After completing the first 8-week exercise training regimen, all subjects will be tested for changes in +G<sub>z</sub> tolerance and associated physiological data, maximal inspiratory and expiratory pressures, peak VO<sub>2</sub>, anaerobic power, and dynamic respiratory muscular strength. Following all post-regimen testing subjects de-train for four weeks. Subjects will, however, continue their 4-cycle SACM centrifuge exposure. After the 4-week period of de-training, all physiologic and physical tests will be collected and the subject will be indoctrinated on the new (crossover design) strength training regimen before the subjects initiate their final 8-week strength training regimen.

**Centrifuge Procedures.** Although peak relaxed +G<sub>z</sub> tolerance will be established and periodically re-assessed for all subjects, the principal criterion by which effectiveness of the exercise conditioning regimens will be evaluated is the duration, in seconds(s), subjects can endure the +4.5 to +7.0 G<sub>z</sub> SACM profiles. For training and data collection runs, the centrifuge will be configured with the upright (13° seatback angle) ACES II-type seat. Foot plates in the centrifuge gondola will be positioned to provide an approximately 120° knee joint angle. Each subject will be properly fitted

with a standard CSU-13 B/P anti-G suit and a standard HGU-55/P helmet with MBU-20/P mask for all centrifuge exposures. Acceleration profiles for pre-and post-experimental data collection consist of:

- a. GOR (1 +G<sub>Z</sub>/s) to relaxed peak +G<sub>Z</sub> with standard anti-G suit on but uninflated;
- b. One cycle of ROR (6 +G<sub>Z</sub>/s) to +5 G<sub>Z</sub> for 15-s (warm-up), followed by 2-min rest;
- c. standard +4.5 to +7.0 G<sub>Z</sub> SACM to fatigue or other end-point criteria [medical monitor decision, self-initiated decision, nausea, 100% peripheral light loss (PLL), 50% central light loss (CLL), or G-induced loss of consciousness (G-LOC)];
- d. 5-min resting recovery in gondola;
- e. repeat SACM profile, step (c); times for both trials will be analyzed separately.

Testing of Aerobic Capacity and Anaerobic Power. Each subject will perform cycle ergometry exercise tests to determine both aerobic capacity and anaerobic power output (Bar-Or 1978) . For both tests, subjects will report to the Exercise Physiology Laboratory (Building 170) dressed in clothing suitable for exercise.

The subject's VO<sub>2</sub> peak (peak aerobic capacity) will be determined while the individual performs an incremental cycling protocol on a cycle ergometer. Every two minutes, the workload will be increased until the subject reaches volitional fatigue. Achievement of VO<sub>2</sub> peak will be determined when at least two of the following criteria are met: leveling of oxygen consumption (VO<sub>2</sub>); respiratory exchange ratio (RER) above 1.15; attainment of predicted maximum heart rate. Gas exchange will be assessed using the Sensormedics 2900z Metabolic Cart for determination of VO<sub>2</sub> and RER values.

The Wingate Anaerobic Test is a 30-s, maximal effort, cycle ergometer task (Bar-Or 1978).

The warm-up for the test consists of 3-5 min of pedaling at a moderate load (@175 Watts), interspersed with two sprints of 3-5 s each. For the test, the subject pedals as fast as possible against a workload ( $.075 \text{ kg} \times \text{kg body weight}^{-1}$ ) which requires a maximal effort for 30-s. Recovery includes 1 min of pedaling against no resistance, followed by 4 min of pedaling against a light-to-moderate resistance (approximately 100 Watts).

Both the  $\text{VO}_2$  peak (aerobic) and Wingate (anaerobic) tests may be terminated because: 1. the subject chooses to stop the test for any reason; 2. the subject reaches volitional fatigue; 3. indications of abnormal cardiovascular response are present; 4. the subject experiences light-headedness, pallor, nausea; or 5. the equipment fails.

**Static Strength Training and Testing Procedures.** The static strength training regimen entails supervised performance of progressively increasing static muscular contractions of the hip and knee extensors (leg press) against fixed foot boards in the modified ACES II seat ("static trainer"). After adjusting the seat to establish a  $120^\circ$  knee joint angle, each subject will perform a general calisthenic warm-up before performing the initial 5-step (150 s) static leg press exercise set. Preparatory to initiating the training regimen, the subject will secure himself herself in the static trainer seat and perform a single 5-s static leg press effort (specific warm-up) at 120 kg, as presented on computer-generated visual prompts. At one minute after completion of the warm-up, the subject will continue to follow computer prompts to initiate and maintain the progressive static leg press exercise training regimen. Subjects will press simultaneously on left and right foot pedals fitted with strain gauge transducers which transmit force signals to a microcomputer. Processed signals then are converted into kilogram (kg) force signals and recorded for subsequent reduction and data analysis. A computer-generated video bar display will indicate the static force the subject produces. The subject's registered effort must not vary by more than  $\pm 5\%$  from the required static effort displayed on the video. The successive, continuous progression of 30-s static efforts (120, 160, 200, 240, and 280 kg) continues until the subject completes 150 s of static leg press effort.



During an initial static strength test, the greatest of the forces generated during three maximal leg press force efforts will be recorded as the static strength value for the leg press. Maximal static efforts will not exceed 5.0 s, and a 2-min resting recovery will be provided between efforts. An additional test of static strength/endurance requires subjects to perform their best static leg press regimen, performed according to computer-generated prompts. As in training, during this strength test, the subject will attempt to maintain force levels of 120, 160, 200, 240, and 280 kg for 30 s at each of the five respective prescribed force levels. The test is completed when the subject completes the entire 150s or (more likely) is unable to maintain the solicited force requirement. A subject's score will be the number of seconds (0-150) he/she is able to sustain the force displayed on the feedback monitor. A standard 10-minute recovery will be required between the maximal leg press strength test and the 150-s sustained force test to allow subjects to recover from the previous test.

General, Dynamic Strength Training and Testing Procedures. Training and testing of abdominal strength/ endurance will be accomplished by performance of trunk flexion exercises ("abdominal crunches"), according to USAF published guidelines (AFP 11-404). Abdominal strength / endurance training consists of three sets of abdominal crunches to fatigue. Self-resisted neck strength exercises ("towel exercises") will be performed by subjects in both exercise groups to reduce the risk of neck injury during centrifuge exposures. However, because of the added risk of neck injury associated with maximal testing, neck strength will not be tested. Except for abdominal crunches, neck exercises, and dumbbell use in training and testing elbow flexion (biceps curl) strength, dynamic strength training and testing will be performed on a multi-station weight stack apparatus (universal Gym). A 6 to 8 repetition warm-up at relatively light weightloads (@50% of 1 RM) for each lifting exercise will be performed by all subjects prior to the first exercise set during training and before initiating heavy lifting during strength tests to determine the 1RM. Exercises consist of leg press, knee extension, knee flexion (leg curl), supine bench press, arm (triceps) pushdown, and lat pull-downs.

Abdominal strength/endurance will be determined as the number of abdominal crunches properly performed in a 60-s set. Arm flexion (biceps) strength will be assessed from 1 RM tests with a barbell (free weights). Other dynamic strength testing consists of determining the 1RM weight load for each of the dynamic exercises performed on the weight stack apparatus. All tests will be preceded by a light resistance warm-up of 6 to 8 repetitions. Next, starting at a weight load the exercise supervisor estimates as approximately 75% of the subject's 1 RM, the subject will attempt single lifts at several progressively greater weight loads until a weight is attempted he/she is unable to lift. After each accomplished lift, the supervisor will increase the weight as the subject rests for 60 s. As the weight load approaches the subject's maximal lift capability, the supervisor will make smaller weight load increases and the subject's rest period will increase to 90. When a weight is attempted which the subject is unable to lift, he/she will rest for 90 s before making a second attempt at the same failure weightload. The highest weight successfully lifted is recorded as the subject's 1RM score. The 1 RM should be determined by not more than the fifth or sixth weight increase. This procedure will be repeated for all dynamic exercises. Strength tests conducted at the end of the 6-week training regimen will be initiated at approximately 1.15 times the subject's most recent 5-7RM training weight load.

Dynamic strength training will be conducted in three sets at 80% of the 1RM (@ 5 to 7RM) weight load. In order to maintain the progressive resistance format of strength training, supervisors will add weight, as appropriate, to the training weight load once a subject demonstrates his/her ability to perform more than eight repetitions (8RM) on the third set of that exercise. Additionally, supervisors will encourage subjects to perform to fatigue on every set of all exercises. A rest interval of 60 seconds will be provided between sets of all dynamic lifting exercises. Exercise supervisors will encourage subject compliance and will record subject attendance and performance (RM workload and repetitions) for all exercises.

Statistical Analysis: The 2 x 2 factorial design incorporates an analysis on variance (ANOVA) with pairwise comparisons to test for significant differences between mean scores of the dependent variables according to the two levels of strength training regimens and two levels (pre and post) of test time. Regression analysis also will be performed to determine if +G<sub>z</sub> SACM endurance times can be predicted from strength scores obtained from subjects prior to each strength training regimen. Additionally, correlation tests (Pearson) will be run to determine if significant relationships exist between means or changes in means of pertinent dependent variables.

### Results, Conclusion

The present investigation is currently in progress thus, the results and thereby the conclusions can not be reported. The duration of this investigation is 22 weeks, thus the project and its results will be forthcoming.

## References

Air Force Pamphlet 11-404, Flying Operations. G-Awareness for Aircrew, 19 August 1994  
Section C-Application.

Bar-Or O. A new anaerobic capacity test-characteristic and applications. Proceedings of the 21st World Congress in Sports Medicine. Sep 7-12, 1978, Brasilia.

Bulbulian R, Crisman RP, Thomas ML, Meyer LG. The effects of strength training and centrifuge exposure on +G<sub>z</sub> tolerance. *Aviat Space Environ. Med.* 1994; 65: 1097-1104.

Burton RR. G-induced loss of consciousness: definition history, current status. *Aviat Space Environ. Med.* 1987; 59; 2-5.

Burton RR. Simulated aerial combat maneuvering tolerance and physical conditioning: current status. *Aviat Space Environ. Med.* 1986; 57; 712-4

Cote R, Tripp L, Jennings T, Karl A, Goodyear C, Wiley R. Effect of inspiratory volume on intrathoracic pressure generated by an L-1 maneuver. *Aviat. Space Environ. Med.* 1986; 57: 1035-8.

Gillingham KK. High G stress and oriental stress: physiological effects of aerial maneuvering. *Aviat. Space Environ. Med.* 1988; 59 (suppl) A10-A20.

Lake FR, Henderson K, Briffa T, Openshaw J, and Musk AW. Upper-limb and lower limb exercise training in patients with chronic airflow obstruction. *Chest* 97: 1077-1082, 1990.

Tesch PA, Hjort H, Balldin UI. Effects of strength training on G tolerance. *Aviat Space Environ. Med.* 1983; 54(8) 691-695.

**Associate did not participate in program.**

**HSP70 Expression in the CNS in Response  
to Exercise and Heat Stress  
in Rats**

**T.J. Walters<sup>1</sup>, M.B. Jones<sup>2</sup>, K. L. Ryan<sup>2</sup>, and P.A. Mason<sup>1</sup>**

**<sup>1</sup>Systems Research Laboratories at  
Armstrong Laboratory, Radiofrequency Radiation Division  
Brooks Air Force Base, TX 78235-5324**

**<sup>2</sup>Department of Biology, Trinity University  
San Antonio, TX 78212**

**Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory**

**Sponsored by:  
Air force Office of Scientific Research  
Bolling AFB Washington DC  
and  
Armstrong Laboratory**

**July 1995**

## Abstract

The expression of the 72-kDa family of heat shock proteins (hsp70) has been studied in a number of tissues in response to exercise. However, hsp70 expression has not been investigated in the brain following exercise. The purpose of this study was to: 1) determine if hsp70 was expressed in the brain in response to acute treadmill (TM) exercise in rats under thermal neutral conditions (24 °C); 2) determine the pattern of hsp70 expression in the brain following TM exercise in a hot environment (34 °C); and 3) compare the pattern of hsp 70 expression in the brain between these groups as well as to sedentary rats exposed to environmental conditions designed to mimic the increase in brain temperature experienced by the two exercising groups. Male Sprague-Dawley rats underwent stereotaxic surgery to implant a guide cannula designed to accommodate a temperature probe used to measure hypothalamic temperature ( $T_{hyp}$ ). Following recovery from surgery, rats were familiarized with TM exercise (13 m/min, 0% grade, 10 min, 3 sessions minimum). Rats were then randomly assigned to one of four treatment groups: 1) neutral/sedentary (NSed); 2) hot/sedentary (HSed); 3) neutral/exercise (NEx); and 4) hot/exercise (HEx). The exercise groups were subjected to 60 minutes of running (21 m/min, 8% grade) at an environmental temperature of 24 °C or 34 °C on a modified TM equipped with environmental temperature controls. HSed rats were placed on the TM (but not run), the environmental temperature as adjusted during testing to mimic the thermal profiles observed in the exercise group. Home cage control rats (Con) were also investigated. Rats were euthanized 6 hr after treatment. Immunohistochemical stained sections were examined using light microscopy and hsp70 expression was quantitated by determining the concentration of positively stained cells in five different brain regions. There was no significant difference between in hsp70 expression among Con, NSed, and NEx groups in any brain region examined. This was in spite of the fact that significant elevations in  $T_{hyp}$  occurred in response to HSed and NEx (mean peak  $T_{hyp}$  =  $40.5 \pm 0.5$  °C). In contrast, the cerebellum of the HEx rats displayed a robust expression of hsp70. From these results it was concluded that hsp70 does not play a role during acute exercise under thermal neutral conditions.

## Introduction

Cells respond to heat and other external stressors with rapid transcription and subsequent preferential translation of a group of highly conserved stress proteins known as heat-shock proteins (hsps). The most prominent stress proteins in eukaryotic cells have molecular weights of approximately 70 kDa and are collectively known as the hsp70 family (see review: Lindquist and Craig, 1988; Moseley, 1993). In rodent cells, hsp73 is present under control conditions and is only slightly heat inducible. Hsp73 might be involved in cellular repair and protection of stressed cells and its function augmented by hsp72, a protein not normally observed under control conditions and considered to be stress induced (Lindquist and Craig, 1988; Moseley, 1993). Another postulated role of hsp73 is to "chaperone" proteins synthesized on free polysomes and then imported into the mitochondrion (Chaing et al., 1989; Hightower et al., 1991), whereas hsp72 binds to unfolded or denatured proteins. The role of hsps beyond this point is unclear. They might restore a protein's tertiary structure or target the protein for degradation (Tytell et al., 1993).

Recent studies in the rat, as well as humans, have demonstrated that hsp70 expression takes place in skeletal muscle, cardiac muscle, liver, and leukocytes in response to acute treadmill exercise (Flanagan et al., 1992; 1995; Locke et al., 1990; Ryan et al., 1991; Salo et al., 1991; Skidmore et al., 1995). hsp70 expression has been demonstrated to take place in the CNS in response to a number of stressors (see review: Lindquist and Craig, 1988; Moseley, 1993; Tytell et al., 1993). However, it has never been examined following acute exercise. Recent studies from our laboratory have shown that hypothalamic temperature ( $T_{hyp}$ ) can increase significantly (40-41 °C) in rats in response to TM exercise (22 m/min, 8% grade) at thermal neutral conditions (22-24 °C) (unpublished observations). We have recently reported that tympanic temperatures ( $T_{tym}$ ) within this range result in immunocytochemically determined hsp70 expression following microwave induced hyperthermia (Mason et al.). It is well known that the central nervous system (CNS) is particularly vulnerable to heat stress (Tytell et al., 1993). This fact, taken together with our previous observations, combined with the punitive roles of hsp70, would suggest that the heat shock response may play an important protective role during TM exercise in the rat CNS in response to conditions commonly employed during exercise studies.

We therefore tested the hypothesis that hsp70 expression would take place in the CNS of rats following acute TM exercise performed at thermal neutral temperatures (24 °C). In addition we also examined hsp70 expression following exertional and nonexertional heating in a hot environment.



## Methods

*Animal Care.* Male Sprague-Dawley rats (350-375 gm) were obtained from the colonies of Charles Rivers (Wilmington, MA). They were individually housed in standard plastic cages (10.5 x 9.0 x 8.0 in) with water available ad libitum and Purina lab chow at a rate of 14 g/d. The light-dark cycle was 12:12 hr (light on 0600 hr) and the room temperature was maintained at 22 °C. At the time of the experiment rats were 4 to 5 mo old.

The protocol used in these experiments was reviewed and approved by the Armstrong Laboratory Animal Care and Use Committee. The animals used in this study were procured, maintained and used in accordance with the Animal Welfare Act and the "Guide for the Care and Use of Laboratory Animals" prepared by the Institute of Laboratory Animal Resources - National Research Council.

*Stereotaxic Surgery.* Rats were anesthetized with a combination of ketamine (60mg/kg; i.p.) and xylazine (10mg/kg; i.p.) and placed on a thermostatically controlled water heating pad set to maintain a  $T_{re}$  of 37 °C. Rats then had a Vialon® guide (Becton Dickinson) stereotaxically implanted into the lateral hypothalamus. The bottom of the guide was sealed and the top of the guide was equipped with a treaded head cap. The cap mated with a tether which contained the thermal probe used to measure  $T_{hyp}$  during experimentation. With the mouthpiece of the stereotaxic instrument set at -3.3, the coordinates for the tip of the guide were 1.8 mm posterior to bregma, 1.5 mm lateral from midline, and 8.3 mm below dura. The guide was held in place with cranioplastic cement (Plastic One, Roanoke, VA) anchored to stainless steel skull screws (0-8).

*TM Familiarization.* Following recovery from surgery (minimum 8 d) rats were familiarized with TM exercise by running on a motorized TM (Columbus Instruments Model: Dual Economy TM) at a speed of 13 m/min (0% grade). Each rat received a minimum of 3 sessions spaced no less than 3 days apart. This protocol has been shown to result in no measurable training adaptations (Cartee and Farrar, 1987). However, to further guard against a possible confounding effect of TM training, rats were not assigned to treatment groups until after TM familiarization was complete, thus all rats had been exposed to TM running.

*Treatment Groups.* Following TM familiarization, rats were assigned to one of three treatment groups: 1) neutral environment/exercise (NEx) were run at environmental temperature ( $T_{env}$ ) of 24 °C (figure 1a); 2)

hot environment/exercise (HEx) were run at an  $T_{env}$  of 34 °C; 3) Hot environment/sedentary (HSed), were placed on the environmentally controlled TM, but did not run. The environmental temperature was manipulated to mimic the a rise in temperature encountered by HEx rats. Home cage control (Con) rats were also included to control for any possible effect of the experimental manipulation of the animals. Positive control rats were produced by microwave induced hyperthermia in which tympanic temperature had exceeded 43 °C. This treatment has been previously shown to result in extensive hsp70 expression in the brain (Mason et al, 1993;1995; Walters et al, 1995).

*Experimental Procedure.* All experiments were carried out between 0730-0930. On the day of the experiment the rat was brought into the laboratory ( $T_{room}$ : 20.6 - 21.5°C) and placed on a plastic cart top. The rat was then instrumented with a rectal probe (YSI 700) which was inserted 7 cm and fixed to the tail with silk surgical tape. The guide cannula, fixed to the rat's skull, was then connected to a tether containing a temperature probe (Vitek mod. 101), once connected, the probe was advanced until it reached the end of the guide. Baseline temperatures were recorded for 10 minutes. The rat was then placed on the treadmill where  $T_{env}$  had been adjusted to the target temperature (see detailed description below), and an 8% grade had been established. In the case of the exercise groups, the rat was placed on the moving TM belt, which was set at a speed of 13 m/min. After 2 minutes the speed was increased to 18 m/min for another 2 minutes. After this warm-up period the TM speed was increased to 21 m/min. HEx rats were not able to complete the entire treatment period at this speed. When they became unable to keep pace with the TM, speed was then reduced to 18 m/min. If a rat was still unable to keep pace, the speed was reduced to 13 m/min, however, when possible, the speed was returned to 21 m/min (figure 1a and b). As stated above, the HSed group was intended to mimic the thermal profile of its respective exercise group. It was therefore necessary to complete the experiment with the Hex group in order to provide this profile (figure 1c). Immediately upon completion of the treatment, the rat was removed from the TM and replaced on the cart top. All treatments lasted for 60 minutes and temperature reported throughout recovery. All temperatures were manually recorded at 5 minute intervals, in addition, using a split screen, the instrument displays and a profile view of each rat on the TM was recorded on video tape for back up and more detailed analysis at a latter time.

*Environmental Control.* One lane of the treadmill was modified as an environmental chamber which was part of a closed system heated with a commercially available heating source. During experimentation, the chamber temperature was monitored continuously and was manually controlled to within  $\pm 1$  °C of the target environmental temperature by turning the heating source on or off while air flow was kept constant.

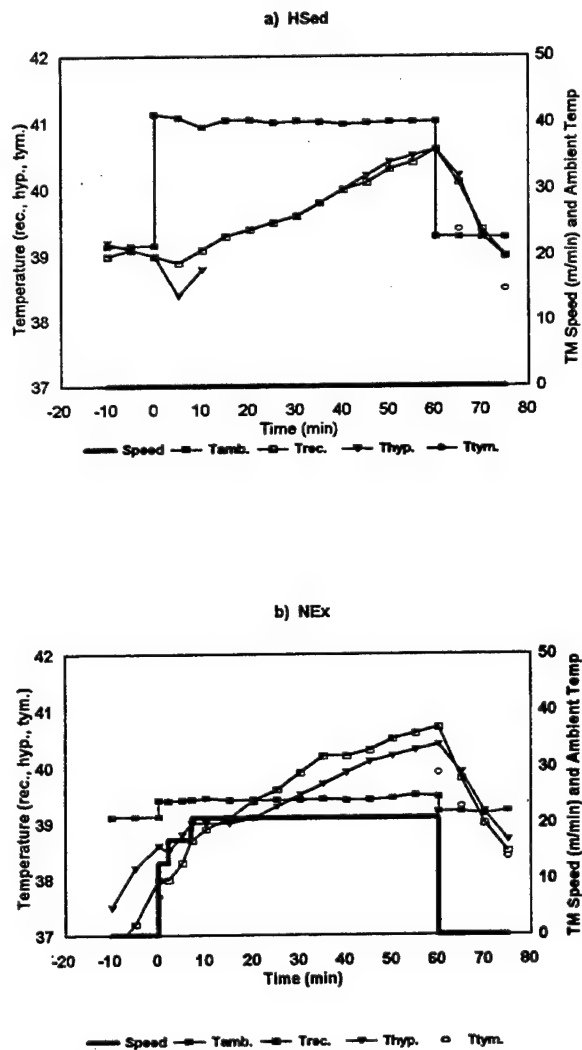
*Immunohistochemistry.* Rats were euthanized with sodium pentobarbital (80 mg/kg, i.p.) 6 hr after termination of microwave exposure. The time of euthanasia was based on previous investigations which demonstrated intense staining 6 hr post HPM exposure (Mason et al., 1993; 1995). When rats were deeply anesthetized and unresponsive to tail pinch, they were perfused intracardially with heparinized saline, followed by a solution of 4% paraformaldehyde/0.1 M phosphate buffer/0.01% thimersol. Brains were removed and stored in this solution for 12 hr followed by soaking in 30% sucrose/0.1 M phosphate buffer until they sank. Brains were then rapidly frozen on dry ice and sliced into 40  $\mu$ m sections on a refrigerated Leica cryostat. Coronal sections corresponding to pages in the rat brain atlas of Paxinos and Watson (1986) were stored in 0.1M phosphate-buffered saline/.01% thimersol until the sections were immunocytochemically stained for hsp-70 (72/73 Kd). The staining method used was compiled from that described by Dragunow and Robertson (1988) and the guidelines produced by Oncogene Science (Manhasset, NY) and Vector Laboratories. The latter was based on the immunoperoxidase procedure devised by Hsu et al. (1981).

Staining consisted of incubating sections in a solution of 1.0% horse serum/0.01 M PBS/0.2% Triton X-100 for 20 min followed by washing with 0.01 M PBS. Sections were then incubated in 1.0 mg/ml hsp-70 primary antibody (mouse monoclonal, hsp 72/73, Oncogene Science) for 24 hr at 4 °C. In negative control sections, incubation in the primary antibody was omitted. Sections were then rinsed in 0.01 M PBS and incubated in biotinylated secondary antibody (mouse IgG, PK-4002, Vector Laboratories) for 60 min at room temperature. After rinsing with 0.01 M PBS, sections were incubated in the Vectastain ABC mixture from the above kit for 60 in at room temperature. After washing with 0.01 M PBS and rinsing with a 0.5% Triton X-100/0.01 M PBS solution, a 0.03% 3,3'diamino-benzidine tetrahydrochloride (Sigma)/0.0015% hydrogen peroxide solution was placed on the sections for 1-3 min. To stop the reaction, sections were placed in distilled water and mounted on gelatin-coated slides. After air drying, sections were dehydrated in 70 and 95% alcohol and 100% 2-propanol for 1 min each. Sections were soaked in AmeriClear (Baxter Scientific) and coverslipped using Permount (Fisher Scientific). Sections (0.30, 1.30, 2.30, 3.30, 4.30, 5.30, 6.30, 7.30, 10.30 and 11.30 mm behind bregma) were visually scanned and photographed using a light microscope (Vanox, Olympus Corp.).

*Statistical Analyses.* Using the Statistica (StatSoft, Tulsa, OK) software, one-way analysis of variance (ANOVA) was conducted on the density of hsp70 expression in each of selected brain regions as a function of experimental group. Statistical significance was considered attained if  $p < 0.05$ .

## Results

The average weight of the experimental rats was 362 grams. hsp70 expression (fig.2) was determined for five separate brain regions. The cerebellum of the HEx rats demonstrated a significant increase in hsp70 expression ( $p=.01$ ). The medial vestibular nucleus, paratrochlear nucleus, paraventricular hypothalamus, and the medial habenula expressed similar hsp70 densities within their respective groups. A positive correlation existed between peak hypothalamic temperature and hsp density. Table 1 demonstrated highest peak temperatures for the HEx rats, followed by Hsed, and then NEx rats. It also expressed the hyperthermic period, defined as the time the animal maintained a hypothalamic temperature greater than 40.5°C.



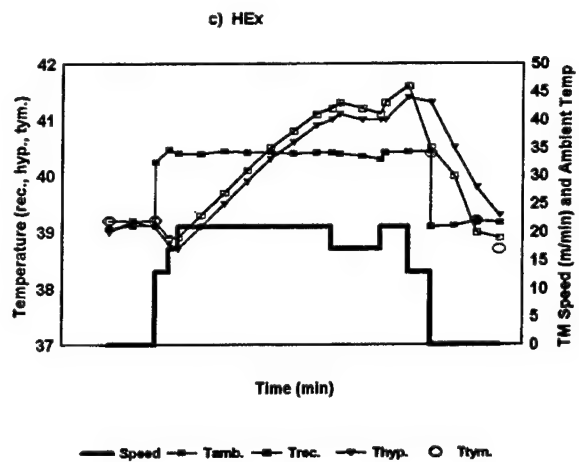
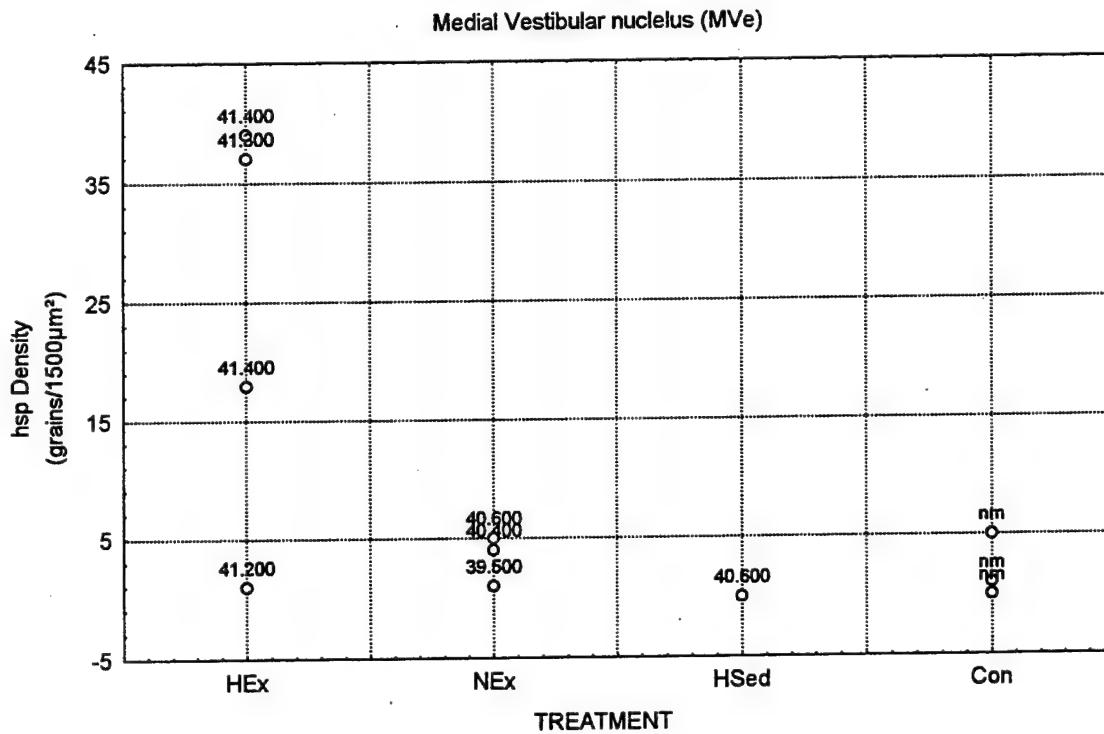
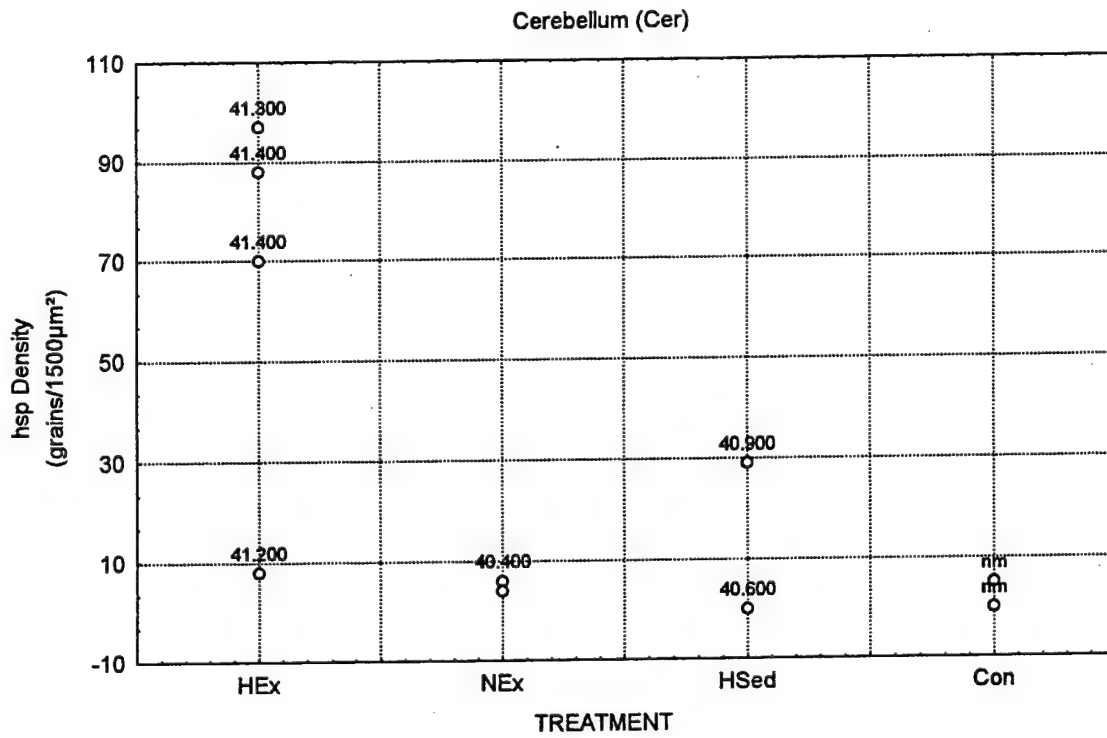
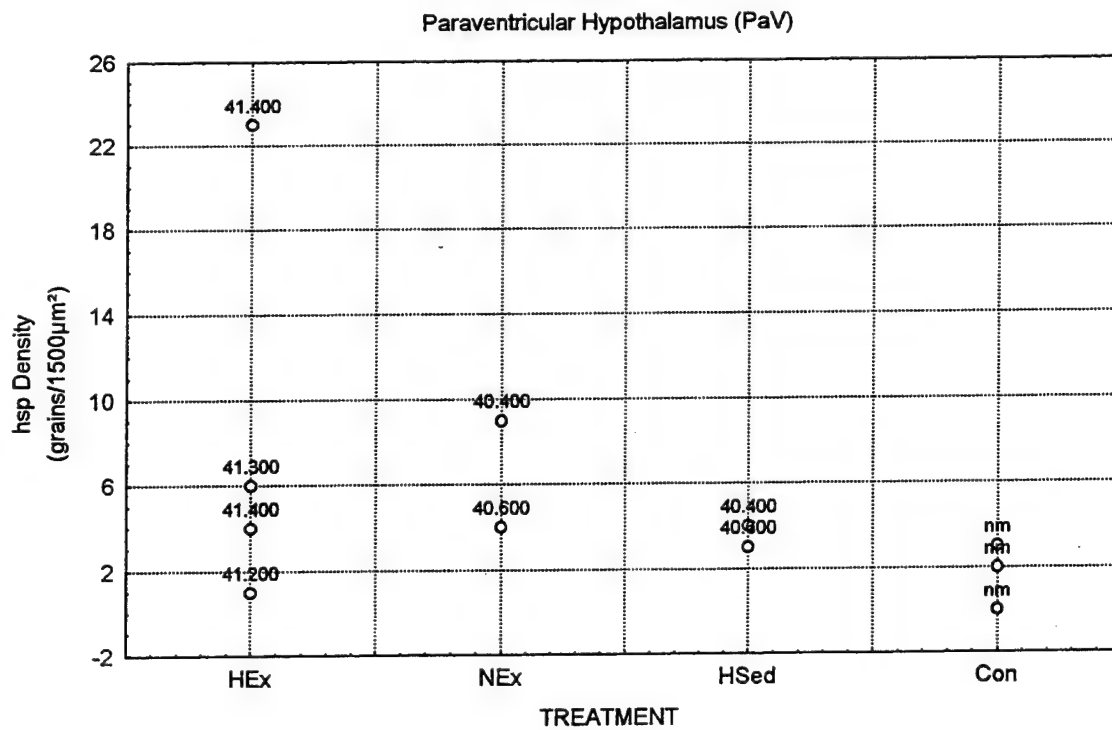
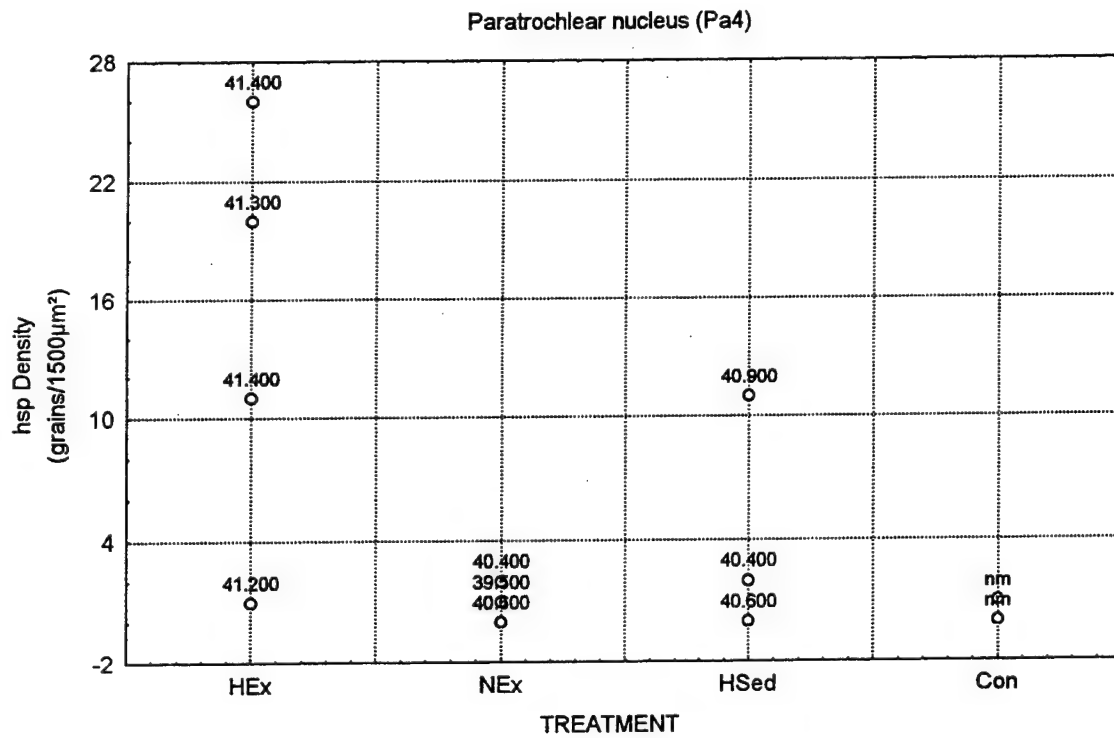
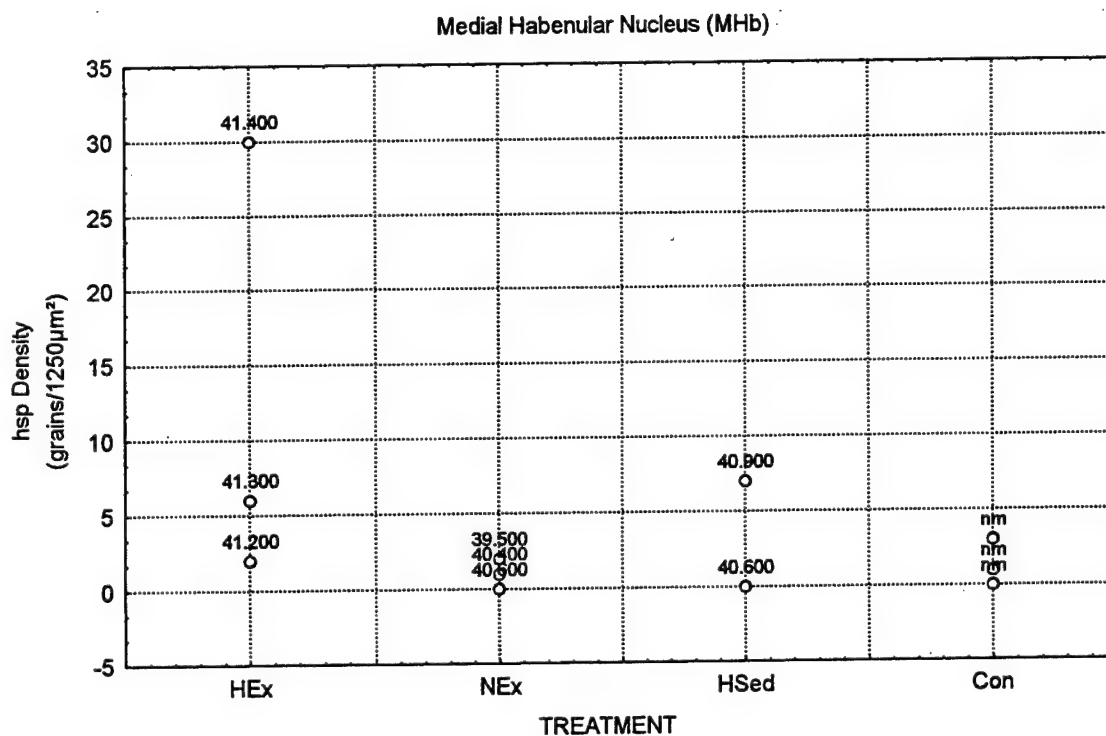


Figure 1a-c. Thermal profiles from rats representing each of the three treatment groups: (a) HSed;  
(b)NEx; (c) HEx







**Figure 2.** Hsp70 density (number of positively stained grains/ $\mu\text{m}^2$ ) from 5 brain regions. The peak hypothalamic temperature is reported for each rat.



TABLE 1. *Thermal and Temporal Data*

Treatment	$T_{\text{rec}}$	$T_{\text{hyp}}$	Time $T_{\text{hyp}} > 40.5^{\circ}\text{C}$
Group	Peak ( $^{\circ}\text{C}$ )	Peak ( $^{\circ}\text{C}$ )	(min)
HSed n=3	40.8	40.6	14.00
NEx n=3	40.3	40.2	2.00
HEx n=4	41.2	41.3	37.25

Values are Means

### Conclusion

The results of this present study do not support our primary hypothesis. The data clearly demonstrates that TM running at  $24^{\circ}\text{C}$  results in a significant increase in  $T_{\text{hyp}}$ , with peaking between  $40 - 41^{\circ}\text{C}$  (Figure 1; Table 1). This observation makes our results more puzzling, as we have demonstrated previously that temperatures in this range do result in significant hsp70 expression, in the brain regions examined, following microwave induced hyperthermia (Mason et al. 1995). As expected, hsp70 expression was correlated with peak hypothalamic temperature. The data suggest that hsp70 expression may be dependent upon a temporal relationship with a threshold temperature of  $40.5^{\circ}\text{C}$  (Table 1), since the hyperthermic period was longer for the HEx treated animals. Interestingly, hsp expression was similar in the single sedentary animal that demonstrated a thermal profile similar to the HEx rats. This suggests that exercise by itself did not influence hsp expression. From our results, we conclude that hsp70 may play a role during acute exercise in a hot environment. But, contrary to our hypothesis, hsp70 is not involved during acute exercise under thermal neutral conditions.

### **Acknowledgements**

The authors would like to extend their gratitude to Ms. Maria Tehrany, Ms. Joanne Doyle, Ms. Janet Kane and Mr. Robert Plottner for there valuable technical assistance.

The animals used in this study were procured, maintained and used in accordance with the Animal Welfare Act and the "Guide for the Care and Use of Laboratory Animals" prepared by the Institute of Laboratory Animal Resources - National Research Council. Views presented are those of the authors and not necessarily those of the U.S. Air Force.

## References

1. Cartee, G.D., and R.P. Farrar. Composition of muscle respiratory capacity and  $VO_2$  max between young and old exercise-trained rats. *J. Appl. Physiol.* 63: 257-261, 1987.
2. Chiang, H.L., S.R. Terleck., C.P. Plant and J.F. Dice. A role for a 70-kilodalton heat shock protein in lysosomal degradation of intracellular proteins. *Science*, 246: 382-385, 1989.
3. Currie, R.W. and F.P. White. Trauma-induced protein in rat tissues: A physiological role for a "heat-shock" protein? *Science*, 214: 72-73, 1981.
4. Dragulow, M. and H.A. Robertson. Localization and induction of c-fos protein-like immunoreactive material in the nuclei of adult mammalian neurons. *Brain Res.*, 440: 252-260, 1988.
5. Flanagan, S.W., A.J. Ryan, C.V. Gisolfi and P.L. Moseley. Effects of heating rate on synthesis of 70K HSP. APS Conference Abstract Issue, Amer. Phys. Soc., 35(4): 213, 1992.
6. Flanagan, S.W., Ryan, A.J., Gisolfi, C.V., Moseley, P.L. Tissue-specific HSP70 response in animals undergoing heat stress. *Am. J. Physiol.*, 268: R28-32, 1995.
7. Hightower, L.E. Heat shock, stress proteins, chaperones and proteotoxicity. *Cell*, 66: 191-197, 1991.
8. Hsu, S.M., L. Raine and H. Fanger. Use of avidin-biotin-peroxidase complex (ABC) in immunoperoxidase techniques: A comparison between ABC and unlabeled antibody (PAP) procedures. *J. Histochem. Cytochem.*, 29: 577-580, 1981.
9. Kirino, T., Y. Tsujita and A. Tamura. Induced tolerance to ischemia in gerbil hippocampal neurons. *J. Cereb. Blood Flow Metab.*, 11: 299-307, 1991.
10. Li, G.C., J.L. Meyer. and G.M. Hahn. Heat-induced protection of mice against thermal death. *Cancer Res.* 43: 5758-5760, 1983.
11. Li, Y., M. Chopp, Y. Yoshida and S.R. Levine. Distribution of 72-kDa heat shock protein in rat brain after hyperthermia. *Acta Neuropathol.*, 84: 94-99, 1992.
12. Lindquist, S. and E.A. Craig. The heat-shock proteins. *Annu. Rev. Genet.*, 22: 631-677, 1988.
13. Locke, M., E.G. Noble and B.G. Atkinson. Exercising mammals synthesize stress proteins. *Am. J. Physiol.* 258 (Cell Physiol., 27): C723-C729, 1990.
14. Locke, M., E.G. Noble and B.G. Atkinson. Inducible isoform of HSP70 is constitutively expressed in a muscle fiber type specific pattern. *Am. J. Physiol.*, 261 (Cell Physiology., 30): C774-C779, 1991.
15. Mason, P.A., T.J. Walters and J.M. Doyle. Heat-shock protein expression in rat brain after microwave exposure. *Soc. Neurosci. Abst.*, 19: Part 1, p.100, 1993.
16. Mason, P.A., T.J. Walters, J.M. Doyle, J.L. Kane, R. Escarciga and K.E. Purdy. Heat-shock protein and GFAP expression after microwave exposure or warm water immersion. *Experimental Biology Abstract*, 1995.

17. Moseley, P.L. Exercise, heat and thermotolerance: Molecular Mechanisms. In: Perspectives in Exercise Science and Sports Medicine: Exercise, Heat and Thermoregulation (Gisolfi, C.V., Lamb, D.R. and Nadel, E.R., eds.) Brown and Benchmark, 305-334, 1993.
18. Paxinos, G. and C. Watson. The Rat Brain in Stereotaxic Coordinates, 2nd edition, Academic Press, Sydney, Australia, 1986.
19. Ryan, A.J., C.V. Gisolfi and P.L. Moseley. Synthesis of 70K stress protein by human leukocytes: Effects of exercise in the heat. *J. Appl. Physiol.*, 70(1): 466-471, 1991.
20. Salo, D.C., C.M. Donovan and K.J. Davies. HSP70 and other possible heat shock or oxidative stress proteins are induced in skeletal muscle, heart and liver during exercise. *Free Radic. Biol. Med.*, 11: 239-246, 1991.
21. Skidmore R, Gutierrez, J.A., Guerriero, V., Kregel, K.C. HSP70 induction during exercise and heat stress in rats: role of internal temperature. *Am. J. Physiol.*, 268:R92-97, 1995.
22. Tytell, M., M.F. Barbe. and I.R. Brown. Stress (heat Shock) protein accumulation in the central nervous system. Its relationship to cell stress and damage. In Seil, F.J. (ed): "Advances in Neurology", New York: Raven Press, 59: 293-303, 1993.
23. Walters, T.J., P.A. Mason, J.L. Kane, J.M. Doyle and R. Escarciga. Heat-shock protein expression in young and old rat brains following microwave-induced heat stress. *Experimental Biology Abstract*, 1995.
24. Welch, W.J. and J.P. Suhan. Cellular and biochemical events in mammalian cells during and after recovery from physiological stress. *J. Cell Biol.*, 103: 2035-2052, 1986.

**WHOLE-BODY CENTER OF GRAVITY AND  
MOMENTS OF INERTIA STUDY**

**M. Virginia Marsh  
Graduate Teaching/Research Assistant  
Department of Engineering Science and Mechanics**

**The University of Alabama  
Box 870278  
Tuscaloosa, AL 35487-0278**

**Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory**

**Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC**

**and**

**Armstrong Laboratory**

**August 1995**

## **WHOLE-BODY CENTER OF GRAVITY AND MOMENTS OF INERTIA STUDY**

**M. Virginia Marsh  
Graduate Teaching/Research Assistant  
Department of Engineering Science and Mechanics  
The University of Alabama**

### **Abstract**

The primary objective of this effort was to conduct precise center of gravity and inertial property measurements on a selected pool of subjects to provide an independent accuracy assessment of the Wright Laboratory Center of Gravity and Inertia Meter (CGIM). Ejection safety is one of the most critical issues for accommodating the Joint Primary Aircraft Training System extended pilot population. Occupant mass, center of gravity (CG,) and mass distribution have a direct effect on ejection accelerations and forces, seat stability and control, harness fit, parachute opening shock, and spinal injury potential. Together, these variables are part of the seat's design limits which could lead to pitch stabilization problems and limb flail. The procedure to measure the whole-body CG and mass moments of inertia (MOI) of human subjects consisted of measuring the properties of a test subject secured within a lightweight chair and frame to provide the properties of the test subject alone. A scale and moment table were used to calculate the CG, a Space Electronics Mass Properties Instrument (MPI) was used to measure the MOI, and a computer and associated software were used to perform all necessary calculations. This system was used to determine CG and MOI data for the chair alone as well as for the composite system containing the chair and a test subject such as manikins, calibration slugs, or a human. Since the error calculations were within an acceptable tolerance, approval was sought to begin scheduling, preparation for, and actual testing of human subjects.

## **WHOLE-BODY CENTER OF GRAVITY AND MOMENTS OF INERTIA STUDY**

**M. Virginia Marsh**

### **INTRODUCTION**

On 28 April 1993, the Secretary of Defense lifted restrictions preventing women from flying combat aircraft. Shortly thereafter, women were flying ejection-seat-equipped aircraft, and specifications for selection of the Joint Primary Aircraft Training System (JPATS) were changed to accommodate this new expanded population. This prompted Congress to inquire about the safety ramifications of this historic decision. By opening the doors to at least 82% of the female population, the new law also opened doors for males on both the smaller and larger anthropometric extremes for current ejection seat designs. [5]

Ejection safety is one of the first and most critical issues for accommodating the female and expanded male population. All operational ejection seats in service for the Navy and Air Force were designed to accommodate a nude male weight range of approximately 135-220 pounds. The new expanded population stretches the weight range on both the smaller and larger ends to 98-245 pounds. Similarly, sitting height restrictions were loosened to allow shorter people to fly. Ejection seat performance is dependent largely on the occupant's size and weight. More specifically, occupant mass, center of gravity (CG) and mass distribution have a direct effect on ejection accelerations and forces, seat stability and control, harness fit, parachute opening shock, and spinal injury potential. Together, these variables are part of the seat's design limits which could lead to pitch stabilization problems and limb flail injuries. Smaller occupants with less mass and lower moments of inertia (MOI) could experience excessive ejection forces during the seat's initial catapult stroke and possibly tumble erratically as the ejection seat hits the airstream. There is a likelihood then that the small occupant will be out of position for the parachute opening, presenting yet another injury potential. Conversely, occupants at the heavier end of the anthropometric continuum may not have enough catapult force to completely clear the aircraft. [5]

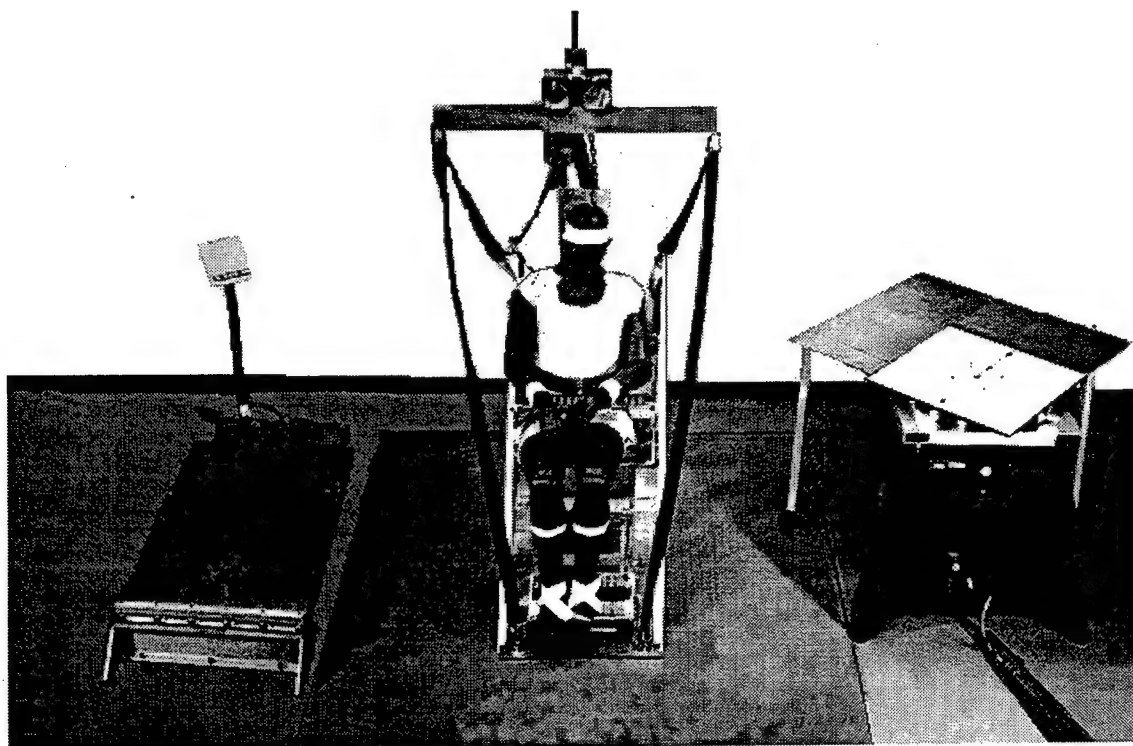
In order to determine the range of whole body CG, an intensive study of human body CG and moments of inertia measurements on a subject population anthropometrically representative of the population accommodated by JPATS or slightly beyond JPATS' boundaries is being conducted. The Vulnerability Assessment Branch in the Biodynamics and Biocommunications Division in the Crew Systems Directorate of Armstrong Laboratory (AL/CFBV) located at Wright-Patterson Air Force Base, Ohio was tasked to conduct precise CG and MOI measurements on a selected pool of subjects to provide an independent

accuracy assessment of the Aircrew Protection Branch in the Vehicle Subsystems Division in the Flight Dynamics Directorate of Wright Laboratory (WL/FIVRA) Center of Gravity/Inertia Meter (CGIM). This joint Navy and Air Force program will measure the CG and MOI of over 30 male and 30 female subjects to predict ejection injury potential. Over 40 anthropometric dimensions will be measured per subject. AL/CFBV and Systems Research Laboratory (SRL) designed and fabricated a lightweight chair assembly for use in the CGIM and AL/CFBV systems. Accurate and repeatable CG and MOI data have been collected for the lightweight chair assembly alone as well as for the composite system containing a test subject such as a manikin, a calibration slug, and a human. The CG and MOI data collected on the CGIM and using AL/CFBV's system will be used to validate the Articulated Total Body (ATB) model's ability to predict net seat and occupant CG. The ATB model simulates human body biomechanics in various dynamic environments, especially aircraft ejection with windblast exposure [4]. The data will be incorporated into the NACES (Navy Aircrew Common Ejection Seat) six-degrees-of-freedom performance model. Similarly, the CG and MOI data will be incorporated into the Wright Laboratory's ACES II ejection seat performance model. This study is a cooperative research effort between the Naval Air System Command, Armstrong Laboratory, Wright Laboratory, and the Naval Air Warfare Center, Aircraft Division (NAWCAD) [5].



## METHODOLOGY

The procedure to measure the whole-body center of gravity (CG) and mass moments of inertia (MOI) of human subjects consists of measuring the properties of a test subject secured within a lightweight chair and frame assembly and subtracting the properties of the lightweight chair and frame to provide the properties of the test subject alone. All measurements are made about the three principal axes: X, Y, and Z. A scale and moment table are used to calculate the CG, a Space Electronics Mass Properties Instrument (MPI) is used to measure the MOI, and a computer and associated software are used to perform all necessary calculations. This experimental configuration is shown in Figure 1.



Floor Scale & Moment Table

Positioning Area

Mass Properties Instrument

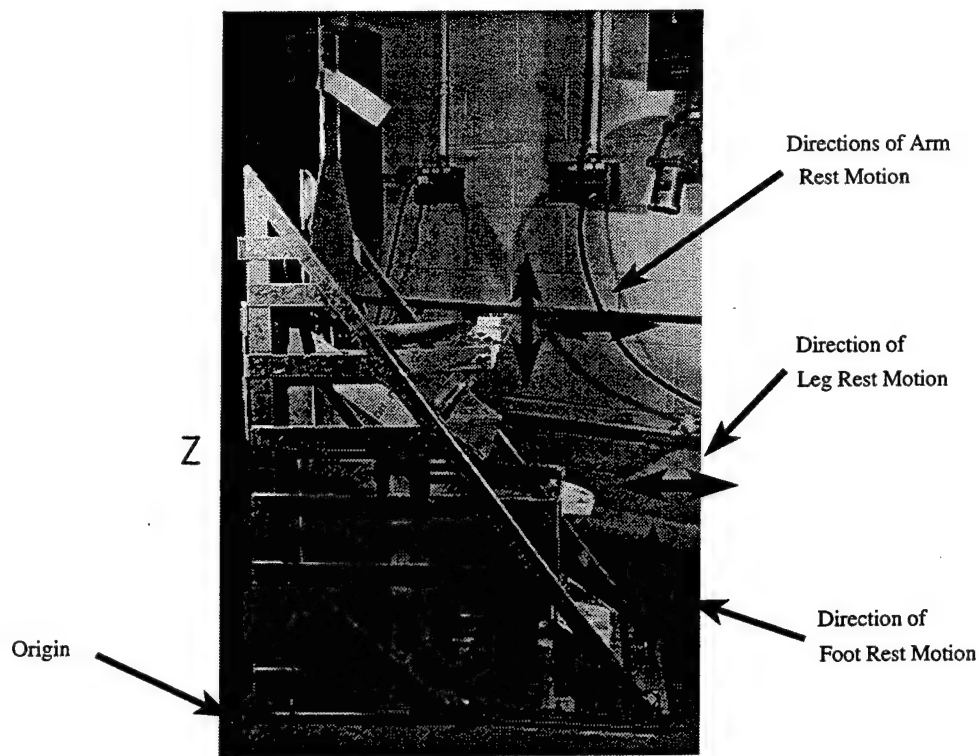
**Figure 1. Experimental Configuration**

### *Description of the Chair Assembly*

The chair assembly consists of a lightweight aluminum chair mounted to a support frame and is shown in Figure 2. The seat incorporates an adjustable seat pan for accommodating various buttock-knee lengths and includes adjustable arm and foot rests. The arm rest has the ability to move both upward and outward to

allow for proper positioning of the elbows at a  $90^\circ$  bend. The leg rest moves parallel to the seat pan to accommodate various buttock-to-popliteal lengths. The foot rest, attached to the leg rest, has the ability to move upward to allow the subject's feet to rest comfortably.

An axis system, based on the structure of the chair and support frame assembly was defined. The origin is defined to be the rear lower RHS corner which is the intersection of the back surface, the bottom surface, and the RHS surface of the support frame. The Z-axis originates at the origin and is directed upward, parallel to the back surface of the support frame. The Y-axis, defined as a vector which is normal to the Z-axis, originates at the origin and is directed to the left. The X-axis originates at the origin and is directed forward, parallel to the bottom surface of the support frame.

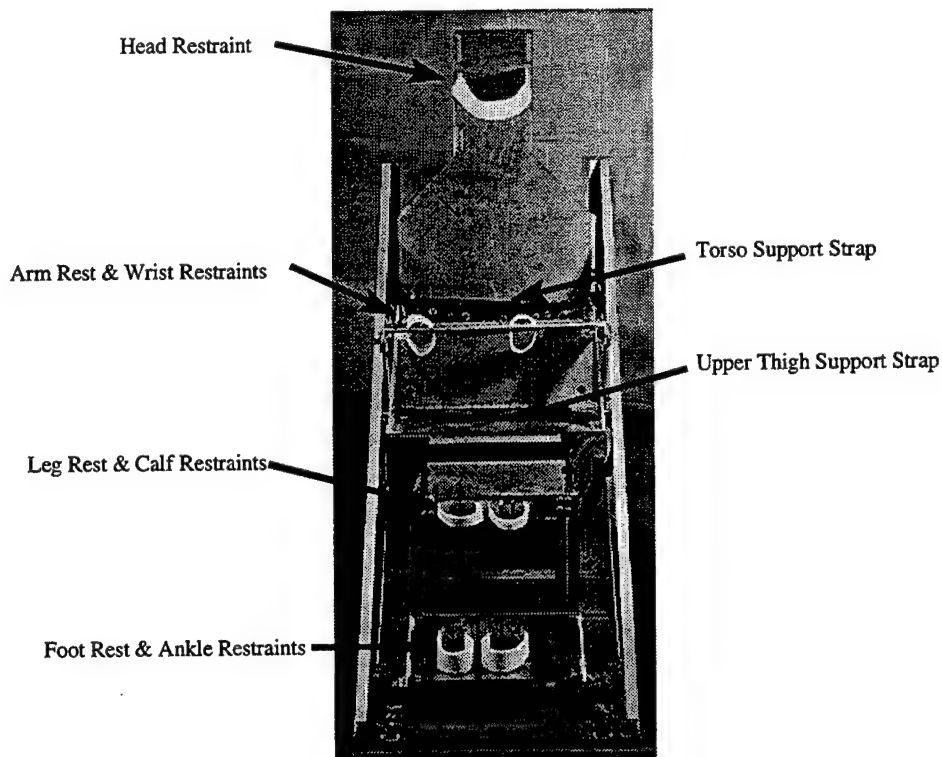


**Figure 2. Lightweight Chair and Frame Assembly**

#### ***Positioning of Individual in Chair***

The first step in the measurement procedure is to secure the subject within the seat. For testing, the seat will be adjusted so that the subject will sit upright against the back rest with the knees bent at a  $90^\circ$  angle. The foot rest will be positioned so that the feet rest comfortably on the foot rest. The legs will be restrained

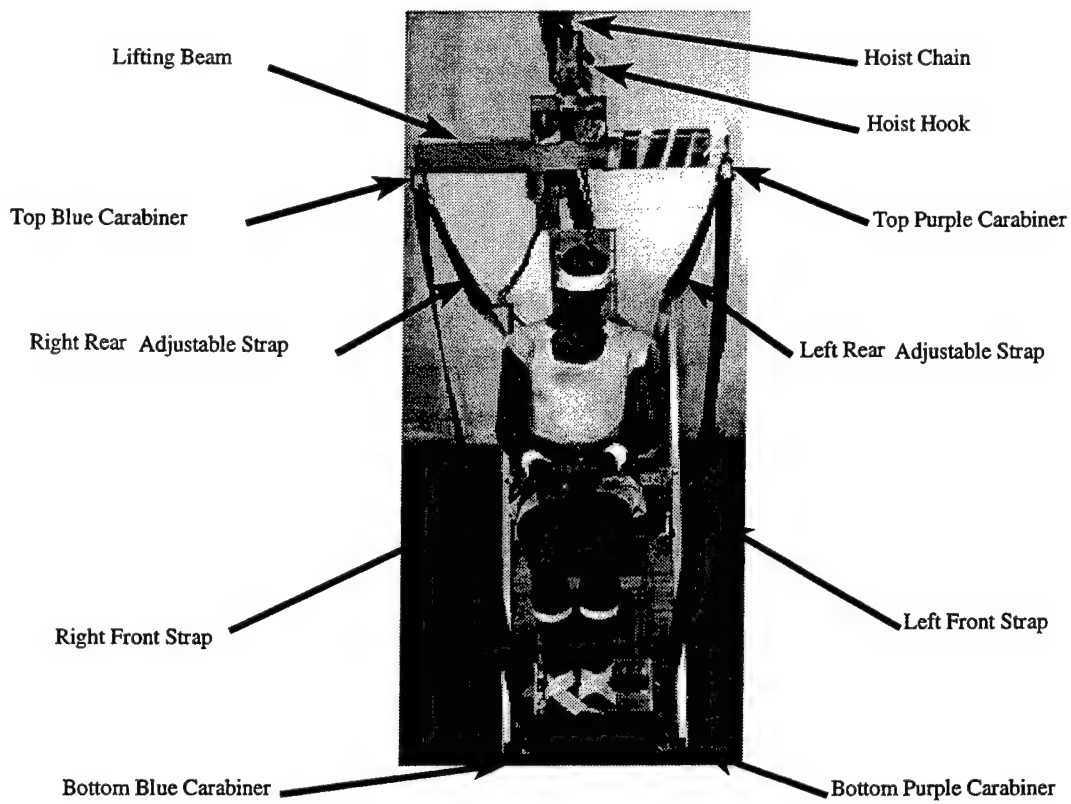
using Velcro straps around the calves and ankles. Rectangular pieces of dense foam will be cut to fill any gaps between the subject's torso and the torso strap attachment points on the seat. The torso strap will be tightened with the foam in place to secure the upper torso within the seat. The arm rest will be positioned such that there is a 90° bend at the elbow. The forearms will be secured to the arm rest using Velcro. Velcro will be used to hold the head securely to the foam head rest. Figure 3 illustrates the seat and restraint system.



**Figure 3. Seat and Restraint System**

#### ***Leveling and Transporting of Chair Assembly***

When the subject has been secured within the seat, lifting straps, attached to the seat will be connected to the hoist hook/lifting beam assembly, as shown in Figure 4. For consistency, the blue carabiner (spring-loaded fastener) should be attached to the right-hand side (RHS) strap and the purple carabiner should be attached to the left-hand side (LHS). Tethering lines will be attached to two sides of the seat and held by test personnel to stabilize the seat during transport of the seat and subject. Positioning of the chair assembly will take place in the area located between the floor scale and moment table and the mass properties instrument, as shown in Figure 1. The chair will be positioned upright, on the back surface and on the Right Hand Side.

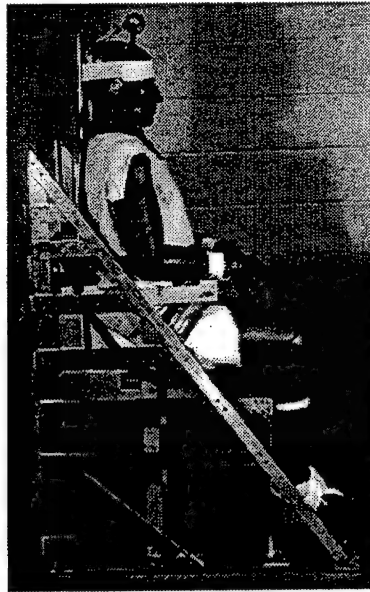


**Figure 4. Subject in Seat with Lifting Straps Installed**

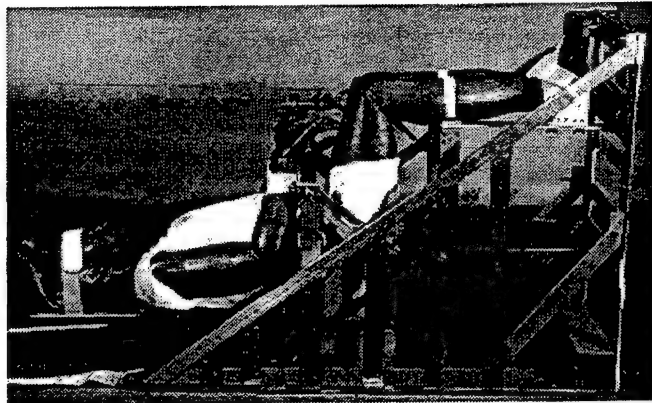
To reduce the amount of discomfort the test subject will experience and to aid in the placement of the seat on the test equipment, it is desired to maintain a nearly level chair assembly during transport. The seat will be raised between measurements to the minimum height which is adequate for repositioning the seat and performing the measurements required by the experiment. The seat will be raised to two different heights: (1) approximately 8" for the floor scale and moment table and (2) approximately 38" for the mass properties instrument.

#### ***Positioning of Chair Assembly***

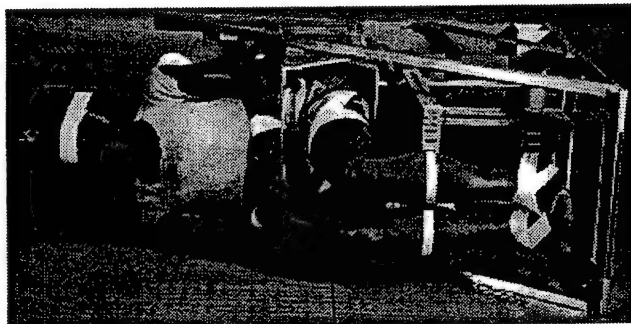
In order to perform measurements of the CG and MOI about the three principal axes, the chair assembly must be positioned in three configurations: upright, on the back surface, and on the RHS. These three configurations are shown in Figure 5.



**Figure 5a. Upright Position**



**Figure 5b. Back Surface**



**Figure 5c. Right Hand Side**

### *Determination of Center of Gravity*

The CG is determined with the use of a scale and moment table assembly as shown in Figure 1. The moment table is a rigid aluminum plate supported by two blades with their edges parallel to each other and separated by a known distance. An aluminum chock is secured to the top of the plate directly above one of the blades. During testing, the chock side of the table is placed on an adjustable stand and the other on the electronic scale.

Prior to testing, the stand is adjusted until the table is level and the scale is tared to read zero.

For testing, the test subject and seat composite is placed on the table with one edge against the chock. The y-z surface must be placed vertically against the chock to measure the moment about the x-axis. The z-x surface is placed against the chock for the y-axis measurement and the x-y surface is placed against the chock for the z-axis measurement. These configurations are illustrated in Figure 6.

The force of the first moment of the test subject and chair assembly along each axis is determined directly from the electronic scale reading. With the mass of the test subject and chair assembly known and the blade-to-blade horizontal separation known, the position of the composite center of mass is calculated by a summation of moments about the chock edge as shown by,

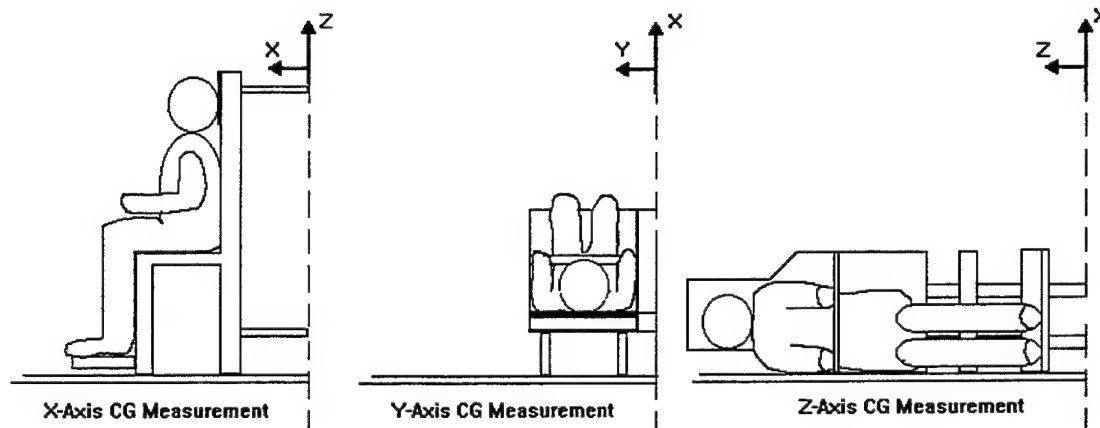
$$F_S R_S = F_{CG} X_{CG}$$

where,  $F_S$  = scale reading of subject and seat on moment table

$F_{CG}$  = weight of subject and seat

$R_S$  = known blade to blade separation distance

$X_{CG}$  = CG coordinate of subject and seat with respect to seat edge in contact with chock.



**Figure 6. Proper Placement of Subject on Scale and Moment Table Assembly for CG Measurement**

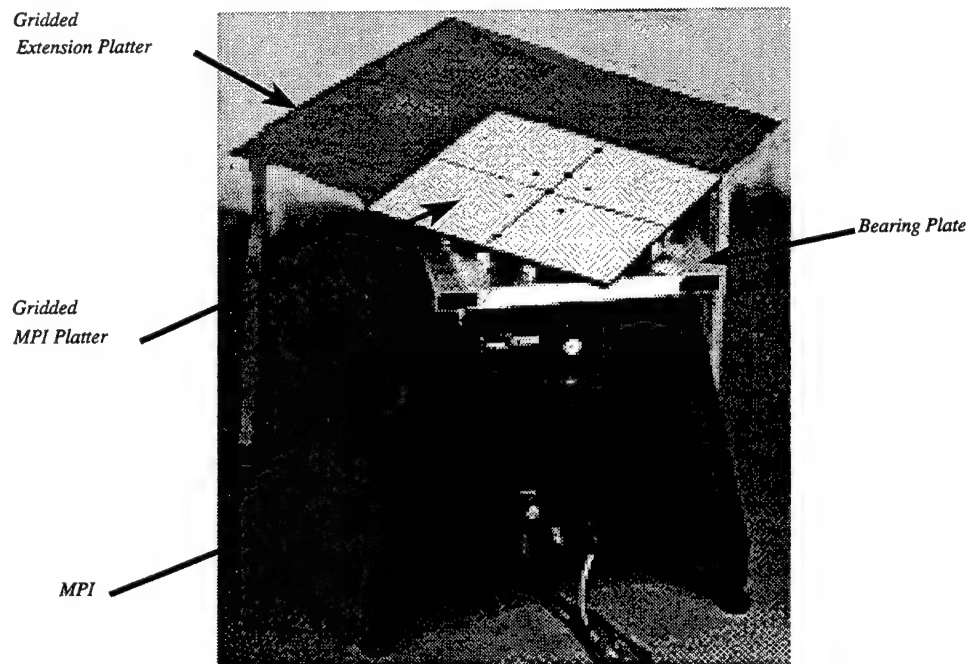
#### ***Determination of Moments of Inertia***

Moments of inertia,  $I$ , of the test object secured within the support fixture are measured using a Space Electronics mass properties instrument. This device, shown in Figure 7, operates on a torsional pendulum technique which uses the fact that the moment of inertia is directly proportional to the square of the period of the rotational oscillation,  $T$ , as shown by,

$$I=KT^2$$

and where  $K$  is the proportionality constant which is a function of the torsional stiffness and length of the torsion rod.

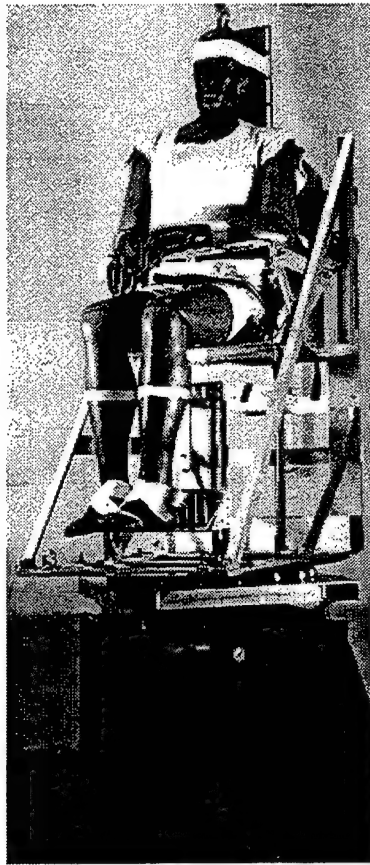
The mass properties instrument is used for measuring objects weighing between 10 and 450 pounds. It employs a torsional flexure technique where the pendulum housed within the instrument is an inverted pendulum composed of a test platter mounted on top of a torsion rod. The period of oscillation is measured by using a photocell device. The motion of the pendulum is strictly rotational and about the centroid of the platter. To ensure this motion, the platter rides on a spherical gas bearing.



**Figure 7. Mass Properties Instrument (MPI)**

The mass properties instrument measures moments of inertia of objects about the pendulum axis. In order to resolve these moments about the center of mass of the test object, the exact location of the center of mass with respect to the pendulum axis must be known. The center of mass coordinates simply designate where to position the composite on the inertial table so that the pendulum axis passes through the composite center of mass. Figure 8 shows the chair assembly and manikin mounted on the mass properties instrument in the configuration necessary for making a measurement about the Z-axis.





**Figure 8. Test Subject Mounted on MPI Platter**

***Measurement of Chair Assembly Without Test Subject***

The whole-body CG and MOI of the test subject will be determined by subtracting the contribution of the lightweight chair and frame from the measured properties of the chair and test subject composite. In order to accomplish this, the properties of the chair and support frame, in the configuration for each subject, need to be measured. The calf, foot, and arm rests will be adjusted to the locations used for each subject and tightened into place.

## RESULTS

The injury potential for individuals subjected to impact is often determined by the mass, center of gravity, and moments of inertia of the various body segments. The STandard Automated Mass Properties (STAMP) Measurement System utilized a digital scale, fulcrum balance, and an inverted torsional pendulum to determine mass, center of gravity, and moments of inertia, respectively [6]. This system was used to determine CG and MOI data for the chair alone as well as for the composite system consisting of the chair assembly with a test subject such as a manikin, a calibration slug, or a human.

### *Chair Assembly*

The STAMP Measurement System provides a time-efficient means of accurately measuring these properties by measuring a test object within a support fixture and subtracting the properties of the support fixture [3]. To obtain support fixture data, the chair assembly was measured to determine its CG and MOI properties. These data were stored to eliminate repeated CG and MOI testing of the support fixture.

### *Chair Assembly With Manikin*

The next phase of testing involved using manikins as test subjects. This was determined by performing CG and MOI measurements of the composite seat and manikin and then subtracting the properties of the chair assembly. The test procedure was slightly altered with the use of adjustable straps to account for the modified CG of the seat and test subject combination. Minor adjustments were required to determine optimal positioning of the equipment and to ease maneuverability of the chair assembly.

### *Validation Using Calibration Slugs*

Experiments were performed with simple calibrated slugs of known mass to determine the accuracy of the measurement process. Validation of the procedure, developed software, and the measurement devices was necessary prior to human subject testing. One means of validation was by measuring CG and MOI values for a 150 lb. calibration slug. This rectangular slug was selected because its CG and MOI values were easily calculated. CG and MOI values for the chair assembly and the calibration slug were measured independently. The results of this phase of the validation procedure are shown in Table One. This table displays an analysis of three repeated measurements. Percent error values were calculated for Calculated Data vs. Measured Data. In block tests 1-3, the slug was mounted in the chair assembly, CG and MOI values were measured, and then the chair properties were subtracted. The test of the block alone was a measure of the block free-standing, not mounted in any holding apparatus.

TABLE ONE. ACCURACY AND REPEATABILITY STUDY USING 150 LB. RECTANGULAR SLUG

Calculated Data								
		(lbs)		(in.)			(lb.-in <sup>2</sup> )	
	Weight:	150.1	CGx:	1.782		MOI X:	9813.377	
			CGy:	2.716		MOI Y:	9603.18	
			CGz:	13.739		MOI Z:	527.961	
Measured Data:								
Block Test #1		(lbs)		(in.)	(% Error)		(lb.-in <sup>2</sup> )	(% Error)
	Weight:	149.7	CGx:	1.936	8.64%	MOI X:	10075.43	2.67%
			CGy:	2.888	6.33%	MOI Y:	9837.942	2.44%
			CGz:	13.7	0.28%	MOI Z:	551.411	4.44%
Block Test #2	Weight:	150.1	CGx:	1.792	0.56%	MOI X:	9782.456	0.32%
			CGy:	2.749	1.22%	MOI Y:	9447.368	1.62%
			CGz:	13.601	1.00%	MOI Z:	456.157	13.6%
Block Test #3	Weight:	150.3	CGx:	1.781	0.06%	MOI X:	10155.27	3.48%
			CGy:	2.712	0.15%	MOI Y:	9889.211	2.98%
			CGz:	13.4	2.47%	MOI Z:	419.963	20.46%
Mean(s):	Weight:	150.03	CGx:	1.8363	3.03%	MOI X:	10004.38	1.95%
			CGy:	2.783	2.47%	MOI Y:	9724.84	1.27%
			CGz:	13.567	1.23%	MOI Z:	475.8437	9.87%
Std Deviation(s):	Weight:	0.24944	CGx:	0.0706		MOI X:	160.2757	
			CGy:	0.0591		MOI Y:	197.3158	
			CGz:	0.1248		MOI Z:	55.43956	
		(lbs)		(in.)	(% Error)		(lb.-in <sup>2</sup> )	
Block Alone:	Weight:	150.1	CGx:	1.789	0.39%	MOI X:	9844.797	0.32%
			CGy:	2.728	0.44%	MOI Y:	9661.972	0.61%
			CGz:	13.762	0.17%	MOI Z:	532.744	0.91%

The largest errors are present in the  $I_z$  moments. This is the smallest moment that is measured for the block, and differences in these values result in larger percentage differences than for larger moments.

A 150 lb. wood and concrete calibration slug was also manufactured so that the three principal MOI values would be of equivalent orders of magnitude. With this slug, accuracy and repeatability tests were also conducted. The attained data were analyzed with percent error values of less than 3% computed when the block alone was measured vs. when the block was measured in the chair assembly. Three repeated CG and MOI measurements were performed using this slug. After each test, the slug was removed from the chair assembly and then repositioned. Replacement of the slug in the same configuration was necessary to obtain consistent CG and MOI results. Straps were used to secure the block to prevent movement during the test procedure. The results found using the wood and concrete slug are shown in Table Two. This table displays an analyses of three repeated measurements of CG and MOI. Percent error values were calculated for the block Measured Alone data vs. the block Measured in Chair data. During block tests 1-3, the slug was mounted in the chair assembly, CG and MOI values were measured, and then the chair properties were subtracted. The first test was a measure of the block free-standing, not mounted in any holding apparatus.

TABLE TWO. ACCURACY AND REPEATABILITY USING 150 LB. WOOD AND CONCRETE  
SLUG

Block Alone Data							
		(lbs)		(in.)			(lb.-in <sup>2</sup> )
	Weight:	148.63	CGx:	5.86		MOI X:	4688.843
			CGy:	6.47		MOI Y:	4287.594
			CGz:	8.42		MOI Z:	3027.045
Measured Data:							
Block Test #1		(lbs)		(in.)	(% Error)		(lb.-in <sup>2</sup> )
	Weight:	148.77	CGx:	5.77	1.54%	MOI X:	4778.984
			CGy:	6.47	0.00%	MOI Y:	4249.461
			CGz:	8.49	0.83%	MOI Z:	2982.301
Block Test #2	Weight:	148.57	CGx:	5.73	2.22%	MOI X:	4781.414
			CGy:	6.45	0.30%	MOI Y:	4261.249
			CGz:	8.53	1.31%	MOI Z:	2978.565
Block Test #3	Weight:	148.68	CGx:	5.69	2.90%	MOI X:	4747.001
			CGy:	6.42	0.77%	MOI Y:	4176.47
			CGz:	8.45	0.35%	MOI Z:	2982.961
Mean(s):	Weight:	148.673	CGx:	5.73	2.22%	MOI X:	4769.133
			CGy:	6.45	0.31%	MOI Y:	4229.06
			CGz:	8.49	0.83%	MOI Z:	2981.276
Std	Weight:	0.081785	CGx:	0.03266		MOI X:	15.6811
Deviation(s):			CGy:	0.009428		MOI Y:	37.49685
			CGz:	0.03266		MOI Z:	1.935577

#### ***Approval from Human Use Review Committee***

Permission from the Human Use Review Committee (HURC) was necessary before testing using human subjects could occur. Reviews were conducted by HURC to evaluate the safety of the test procedure and the experimental configuration. These members made recommendations for modifications to the test environment to ensure the safety of the people involved in the procedure. Upon the implementation of their recommendations, permission was granted to use human test subjects for this study. Prior to actual testing, a human test subject was positioned in the chair assembly and raised to a minimal level to allow valuable feedback to be attained in areas regarding stability, safety, and comfort of the chair assembly. These suggestions were implemented to provide a more stable chair assembly so that error would not be introduced. A miniaturized version of the test plan was formatted into a checklist and submitted to HURC as well as with the finalized test plan. The checklist will serve as a guide to all personnel involved in this testing procedure to ensure consistency in the measurements, efficiency in the procedure, and minimization of discomfort to the test subject. HURC was present when the first human subject was tested.

#### ***Chair Assembly with Human Test Subject***

The final accuracy and repeatability study was performed using a human test subject. Since CG and MOI values for the human are unknown, only values of mean, standard deviation, and coefficient of variance were determined from the attained data. The coefficients of variance are the standard deviation divided by the mean and were less than 2%. Values determined from the human whole-body CG and MOI measurement are shown in Table Three. This table displays weight, CG, and MOI values when a human subject was measured three times, in as many days. The purpose of these data was to establish reliability/repeatability/accuracy of replacement of a human in the chair assembly and the effects of this replacement on the CG and MOI values.

**TABLE THREE. ACCURACY AND REPEATABILITY STUDY USING HUMAN TEST  
SUBJECT**

<i>Measured Data:</i>						
<i>Subject Measured in Chair #1</i>		<i>(lb.)</i>		<i>(in.)</i>		<i>(lb-in<sup>2</sup>)</i>
	<i>Weight:</i>	<i>138.00</i>	<i>CGx:</i>	<i>14.474</i>	<i>MOI X:</i>	<i>19968.902</i>
			<i>CGy:</i>	<i>11.474</i>	<i>MOI Y:</i>	<i>22623.441</i>
			<i>CGz:</i>	<i>34.224</i>	<i>MOI Z:</i>	<i>8307.206</i>
<i>Subject Measured in Chair #2</i>	<i>Weight:</i>	<i>138.00</i>	<i>CGx:</i>	<i>14.408</i>	<i>MOI X:</i>	<i>19435.949</i>
			<i>CGy:</i>	<i>11.441</i>	<i>MOI Y:</i>	<i>22512.316</i>
			<i>CGz:</i>	<i>34.059</i>	<i>MOI Z:</i>	<i>8493.059</i>
<i>Subject Measured in Chair #3</i>	<i>Weight:</i>	<i>138.90</i>	<i>CGx:</i>	<i>14.413</i>	<i>MOI X:</i>	<i>19448.625</i>
			<i>CGy:</i>	<i>11.465</i>	<i>MOI Y:</i>	<i>22413.826</i>
			<i>CGz:</i>	<i>33.969</i>	<i>MOI Z:</i>	<i>8574.287</i>
<i>Mean(s):</i>	<i>Weight:</i>	<i>138.30</i>	<i>CGx:</i>	<i>14.431</i>	<i>MOI X:</i>	<i>19617.825</i>
			<i>CGy:</i>	<i>11.46</i>	<i>MOI Y:</i>	<i>22516.528</i>
			<i>CGz:</i>	<i>34.084</i>	<i>MOI Z:</i>	<i>8458.184</i>
<i>Std Deviation:</i>	<i>Weight:</i>	<i>0.0424464</i>	<i>CGx:</i>	<i>0.030</i>	<i>MOI X:</i>	<i>248.30262</i>
			<i>CGy:</i>	<i>0.0139</i>	<i>MOI Y:</i>	<i>85.62677</i>
			<i>CGz:</i>	<i>0.10559</i>	<i>MOI Z:</i>	<i>111.78928</i>
<i>Coefficient of Variance:</i>	<i>Weight:</i>	<i>0.31%</i>	<i>CGx:</i>	<i>0.21%</i>	<i>MOI X:</i>	<i>1.27%</i>
			<i>CGy:</i>	<i>0.12%</i>	<i>MOI Y:</i>	<i>0.12%</i>
			<i>CGz:</i>	<i>0.31%</i>	<i>MOI Z:</i>	<i>0.31%</i>

## CONCLUSIONS

Occupant mass, CG, and mass distribution have an effect on ejection safety. All operational seats in service for the Navy and Air Force were designed to accommodate a range of male weights, but the newly expanded pilot population extends the range in both the smaller and larger body size directions. Little is known of the effects of smaller and larger body sizes on ejection safety. With the increasing occurrence of smaller and larger pilots, reducing the risk associated with ejection is imperative. The results of this study will be valuable in validating the accuracy of the WL database of human body CG and MOI measurements. Since the error calculations were within an acceptable tolerance, approval was sought to begin scheduling, preparation for, and actual testing of human test subjects. Subjects selected for participation in the study will be anthropometrically representative of the Joint Primary Aircraft Training System (JPATS) extended pilot population or slightly beyond JPATS boundaries. CG and MOI measurements will be performed twice on each subject unless there are significant variations in the two measurements, which would necessitate a third measurement and prompt investigation for sources of error. Several anthropometric measurements will be made using anthropometers to ensure that the test subjects are placed and replaced in approximately the same position in the chair assembly. Testing of a minimum of six human subjects will be performed to yield comparison with the Wright Laboratory (WL/FIVRA) Center of Gravity and Inertia Meter (CGIM) System results. The CGIM will be used to predict net seat and occupant center of gravity and moments of inertia for the full complement of 60 subjects.

## REFERENCES

1. Albery, C.A., Obert, A., Whitestone, J., Lephart, S.A., "The Automated Mass Properties Measurement System Testing Procedure," Harry G. Armstrong Medical Research Laboratory, Report No. AAMRL-TR-89-300, Wright Patterson Air Force Base, Ohio, October 1989.
2. Boynton, R., "A High Accuracy Instrument for Measuring Moment of Inertia and Center of Gravity," SAWE Paper No. 1511A.
3. Miller, D.J., Spittle, E.K., Albery, C.B., "Automated Procedure for Measurement of Mass Properties", Proceedings of the 29th Annual SAFE Conference, November 11-13, 1991.
4. Obergefell, L.A., Gardner, T.R., Kaleps, I., Fleck, J.T., "Articulated Total Body Model Enhancements, Volume 2: User's Guide" AAMRL-TR-88-043, January 1988.
5. Rasmussen, R.R., Jr., "Test Plan: Human Center of Gravity and Mass Moments", Work Unit 72313601, January 1995.



COMPARATIVE STUDY BETWEEN PULSED-FIELD GEL ELECTROPHORESIS AND REPETITIVE-  
SEQUENCE BASED PCR AS METHODS OF TYPING  
*CANDIDA RUGOSA*

Beth Ann McBrearty  
B.S. Biology

University of Scranton  
Scranton, PA 18510

Final Report for:  
Summer Graduate Student Research Program

Armstrong Laboratory

Sponsored by:  
Air Force Office Of Scientific Research  
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1995

COMPARATIVE STUDY BETWEEN PULSED-FIELD GEL ELECTROPHORESIS AND REPETITIVE-SEQUENCE BASED PCR AS METHODS OF TYPING *CANDIDA RUGOSA*

Beth Ann McBrearty B.S.  
Department of Biology  
University of Scranton

ABSTRACT

Many *Candida spp.* are medically important, life-threatening pathogens found in immunocompromised individuals and are highly prevalent in the healthy population as well(3). The goal of this comparison is to determine the best method of typing which produces the fastest, most reliable, and most reproducible results. This goal was obtained by evaluating the effectiveness of two methods of DNA fingerprinting, pulsed-field gel electrophoresis and repetitive-sequence based PCR analysis, which were performed on 17 clinical isolates of *Candida rugosa*. The pulsed-field gel electrophoresis in cooperation with a NotI restriction enzyme digestion was performed using the GenePath system (BioRad Foster City, CA).

COMPARATIVE STUDY BETWEEN PULSED-FIELD GEL ELECTROPHORESIS AND REPETITIVE-SEQUENCE BASED PCR AS METHODS OF TYPING *CANDIDA RUGOSA*

Beth Ann McBrearty

INTRODUCTION

*Candida rugosa* is a more recently identified member of the *Candida* genus, which contains approximately 200 species. This study utilizes samples which have been isolated from human burn patients as well as various animal hosts. Among the *Candida spp.* there exists a wide variety of *Candida* associated diseases. Therefore, there is a need for accurate and rapid identification of all *Candida species*. In this study the effectiveness of two methods, pulsed-field gel electrophoresis and repetitive-sequence based PCR(rep-PCR), will be discussed and evaluated as potential methods for fulfilling this need.

Pulsed-field gel electrophoresis(PFGE) followed by a restriction enzyme digestion is one method that can be employed to generate DNA fingerprints. The PFGE utilized in the study is based on the contour-clamped homogeneous electric field(CHEF)method. The CHEF method achieves separation of DNA molecules by passing electrical currents through a hexagonal array of equally spaced point electrodes(1). This method offers complete uniformity across the electric field as well as significant resolution of bands. The CHEF method in cooperation with restriction enzyme digestion is also designed to produce a banding pattern which consists of at least 10 distinct bands(2). These banding patterns may be further analyzed in epidemiological identification. Despite the capabilities of the PFGE, there are also several disadvantages associated with this method. For instance, the entire process is time consuming and labor intensive(5). Therefore, the procedure would not be conducive in the production of rapid results.

Repetitive sequenced-based PCR utilizes the short, interspersed, repeated

elements contained in the eukaryotic genome as the foundation for its DNA fingerprints. More specifically, oligonucleotide primers, which are complementary to these repeated elements, amplify the regions between the repetitive sequences. Since these regions vary in size, multiple DNA fragments may be produced from one genome. The DNA fingerprint can subsequently be produced by electrophoresing these DNA fragments, which are separated according to molecular weight. This technique has been noted to be less time consuming and to be relatively easy to perform(5). However, the discriminatory power is not as high as the PFGE.

#### METHODOLOGY

*Candida rugosa* clinical samples were provided by Texas Medical Center. The samples were incubated at 30 degrees Celsius on Sabouraud-dextrose agar plates for maintenance of stock cultures.

The method utilized for preparation of intact yeast DNA (plugs) was a modified version of the procedure described in the GenePath Group 4 Reagent Kit(BioRad Foster City, CA). All the reagents necessary to perform this procedure were provided in the GenePath group 4 reagent kit (BioRad, Foster City, CA). Single colony isolates of each sample were inoculated into 3ml of Yeast Peptone Dextrose broth and were grown overnight at 30 degrees Celsius in a shaking incubator. 1.7 mls of the culture was transferred to a sterile microcentrifuge tube, the cells were pelleted at 12,000 RPMs for 5 min and the supernatant was discarded. The cells were resuspended in 100-uls of cell suspension buffer and temporarily stored on ice. 6-uls of lyticase enzyme(25mg/ml) was added to the cell suspension buffer for each sample. The cell-lyticase suspension was transferred to another microcentrifuge tube containing a 100uls of 1.6% CleanCut agarose which was previously aliquoted and stored at 60 degrees Celsius. After the cell-lyticase-agarose suspension

was quickly mixed, the suspension was transferred into a premade mold to form the plug and solidified at 4 degrees Celsius for 15 min. The plugs were removed and placed into their own 2.0 ml microcentrifuge tubes containing 1 ml of the lysis II buffer and 60-uls of lyticase enzyme. The plugs were incubated at 37 degrees Celsius for 3 hrs. After this incubation was complete the liquid was aspirated and the plugs were rinsed with 1 ml of 1x wash buffer. The 1x wash buffer was removed and 1 ml of proteinase k buffer, as well as 40-uls of proteinase k enzyme(25mg/ml), were added to each tube. The tubes were gently mixed by inversion and incubated at 50 degrees Celsius overnight. This proteinase k incubation was repeated once more. After the second proteinase k incubation was completed the plugs were rinsed and stored with 1 ml of 1x wash buffer.

The restriction digestion was carried out with the enzyme NotI(5U/ul) (BioRad, Foster City, CA). To prepare for the digestion the plugs were rinsed with 1 ml of 1x wash buffer along with gentle agitation for 1 hour. The buffer was then aspirated. These steps were repeated twice more. The plugs were then washed with 1ml of 0.1x wash buffer along with gentle agitation for 1 hour. The last wash was with 1 ml of 1x NotI buffer along with gentle agitation for 1 hour. The NotI buffer was aspirated. The digestion was carried out with 7.5 ul of NotI(5U/ul) and 450ul of 1x NotI buffer. The plugs were incubated overnight at 37 degrees Celsius. When the incubation was completed the plugs were washed and stored in 1 ml of 1x wash buffer.

A 1% molecular certified agarose gel was used to make the gel. The running buffer was made with 1% TBE and the program used for electrophoresis was 'Pseudomonas' (Electrophoresis kit Biorad, Foster City, CA).

### RESULTS

Figure 1 shows the first nine clinical isolates which have been digested

with NotI and run on pulsed-field gel electrophoresis. The PFGE system used was the GenePath and the program used was 'Pseudomonas'. The gel provided fingerprints for samples 92-1182, 92-1183, 92-1186, 92-1188, and 92-1190. The rest of the samples failed to work and it was necessary to redo these samples.

Figure 2 displays the second attempt for the samples that did not work in figure 1. The samples that worked included 92-1185 and 92-1189. These samples were also run on the 'Pseudomonas' program for the GenePath PFGE.

Figure 1:

Electrophoresis Profile

Date 7/28/95  
Voltage Pseudomonas  
Program

Well#	Sample
1	<u>S. cerevisiae</u>
2	<u>92-1182</u>
3	<u>92-1183</u>
4	<u>92-1184</u>
5	<u>92-1185</u>
6	<u>92-1186</u>
7	<u>92-1187</u>
8	<u>92-1188</u>
9	<u>92-1189</u>
10	<u>92-1190</u>
11	<u>H. wingei</u>
12	
13	



Figure 2:

Electrophoresis Profile

Date 7/29/95  
Voltage Pseudomonas  
Program

Well#	Sample
1	<u>S. cerevisiae</u>
2	<u>92-1184</u>
3	<u>92-1185</u>
4	<u>92-1187</u>
5	<u>92-1189</u>
6	
7	
8	
9	
10	
11	
12	
13	



## CONCLUSIONS

In order to determine the best method of typing, both PFGE and Rep-PCR were attempted on the clinical isolates of *Candida rugosa*. The work that was performed at Brooks AFB, as part of the AFOSR program, was the PFGE, while the work utilizing the Rep-PCR was carried out at the University of Scranton. There were several obstacles that were necessary to overcome before an accurate assessment of both methods can be determined. For example, the preparation of DNA in the plugs required a sufficient amount of cells in order to generate the amount of DNA necessary for the fingerprints. In addition, another critical factor was complete lysis of the cells. This factor was directly affected by the time of incubation and by the amount of lysing enzyme used in the cell lysis step. Another aspect of the PFGE which needs to be fully explored is the optimization of the NotI restriction enzyme as well as usage of other restriction enzymes in order to generate the best DNA fingerprints. The PFGE work will be continued and collaborated with the Rep-PCR work completed University of Scranton to effectively evaluate the goal of the study.



#### REFERENCES

- (1) Bustamante, C., et al. (1993). Towards a Molecular Description of Pulsed-Field Gel Electrophoresis. Trends in Biotechnology. 11: 23-30
- (2) GenePath Group 4 Regent Kit For Use With *Candida* Instruction Manual. Catalog number 310-0005.
- (3) Murray, P. R., et al. (1995). Manual of Clinical Microbiology. ASM Press, Wash., D.C. 731-32.
- (4) Versalovic, J., et al. (1994). Genomic Fingerprinting of Bacteria Using Repetitive Sequence-Based Polymerase Chain Reaction. Meth. in Mol. and Cell. Biol. 5:25-40.
- (5) Yuk-Fong Liu, P., et al. (1995). Analysis of Clonal Relationships among Isolates of *Shigella sonnei* by Different Molecular Typing Methods. J. Clin. Micro. 33:1779-1783.

The Multivariate Correction For Range Restriction  
Is A Conservative Estimator

Thomas N. Meyer

St. Mary's University  
One Camino Santa Maria  
San Antonio, TX 78228-8573

Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory

Sponsored by:  
Air Force Office of Scientific Research  
Brooks Air Force Base, San Antonio, TX

and

Armstrong Laboratory

August 1995

The Multivariate Correction For Range Restriction  
Is A Conservative Estimator

Thomas N. Meyer  
Graduate Student  
St. Mary's University

Abstract

It is sometimes necessary to study samples that have been range restricted and to estimate correlations in a range restricted sample. These correlations are most often reduced compared to correlations that have been calculated in the population. Of particular interest is the situation in which selection takes place on more than one variable. There are corrections for this multivariate selection and this study investigated them through simulation. The sample size, magnitude of the specified correlation, selection ratio, and number of variables involved in selection were varied. The simulation results indicate that on average the correction was slightly downwardly biased and produced conservative estimates of the population values. Correction procedures are beneficial and should be used to provide better statistical estimates.

## The Multivariate Correction For Range Restriction Is A Conservative Estimator

Thomas N. Meyer

### Introduction

When data are collected in a sample that systematically excluded a proportion of the population, range restriction occurs. The variance of the variables upon which selection took place has been reduced to some value lower than would have been found in the population of applicants. The statistics for the remaining range restricted sample are not the same as those of the unselected population. For example, a test is used to select applicants for a job. The selected applicants are different from the general population. The statistics from the selected applicants will not be the same as in the general population and will differ according to the severity of selection.

In general, range restricted groups provide lower correlations than the population. The more restricted the group, the lower the validity correlation coefficients. Simply stated, less variability in a range of scores will result in a lower validity coefficient.

Fortunately, several range restriction formulas exist. In 1903, Pearson published a single variable solution. This solution was based on three assumptions: linearity of regression form, homoscedasticity of error and normally distributed variables. Linearity refers to the linearity of regression of Y on X. This means that any rate of change of Y is the same over all X values. Homoscedasticity of error means that the conditional variance of Y given X does not depend on X. It may be restated that the error variances are uniform over the complete X distribution. Lawley (1943) discovered that the normality assumption was not necessary.

Although Pearson is credited with the early work, it was Thorndike who brought these formulas into common use. Thorndike characterized three curtailment situations. These were differentiated by which variable produced the range restriction and by which standard deviations were known in both the selected sample and applicant population (Thorndike, 1949). In each case a different correction formula is indicated. These correction formulas, which are referred to as Thorndike Case 1, Case 2, and Case 3, are

only applicable for correcting the correlation when restriction has occurred on either X, Y, or on a third variable.

Case 1, although rarely seen in practice, concerns itself with the correlation between variables 1 and 2.

The restriction in range can be found on variable 1. The standard deviations for both the restricted and the full groups for variable 2 are known. Thorndike (1949) gives the following example. A research test, variable 2, had been given to a random sample of applicants and the standard deviation determined. At a later date, variable 2 was correlated with the selection test, variable 1, for a group who passed the selection test. The standard deviation of the research test (variable 2) would be known in both the unrestricted and restricted groups. The correction formula for Case 1 is

$$rc := \sqrt{1 - \left(\frac{s}{S}\right) \cdot (1 - r_{12}^2)}$$

where rc is the corrected correlation between variables 1 and 2.  $r_{12}$  is the correlation between variables 1 and 2 in the restricted group. The symbol s is the standard deviation of variable 2 in the restricted group and S is the standard deviation of variable 2 in the unrestricted group.

Case 2 also estimates the correlation between variables 1 and 2. In Case 2 the restriction is again on variable 1 however, the standard deviations of variable 1 in both the unrestricted and restricted groups are known. For example, a college uses SAT scores (variable 1) as the basis for admission and then only accepts students with scores above 1400. If the scores of only the accepted group were then correlated with some criteria (variable 2) such as college grade point average, a low correlation will result. However, in this situation the correction formula provides a better statistical estimate. The correction formula is:

$$rc := \frac{r_{12} \cdot \frac{S}{s}}{\sqrt{(1 - r_{12}^2) + r_{12}^2 \cdot \frac{S^2}{s^2}}}$$

where the symbols are the same as in Case 1. Case 2 is important because it is very common in the estimation of test validity, especially in education and personnel selection.

The last situation where range restriction can occur is Case 3. This case is very common. Once again, the emphasis is upon the correlation between variables 1 and 2 but in this case the restriction can be found on a third variable which will indirectly affect the variance of the other two variables. In this situation the standard deviations of the third variable, for both the restricted and unrestricted group, must be available. A case three example was given by Ree, Carretta, Earles, & Albert (1994). Suppose a test were administered to a group of applicants, some of whom were selected for employment on the basis of that test, and its variance is known in both the restricted sample and the unrestricted sample. Let scores for Test 2 and a job performance criterion Y be obtained on this restricted sample and all three bivariate correlations calculated. If the correlation between tests 1 and 2 is not zero, then direct selection on 1 will lead to indirect (often called incidental) selection on 2. In this case the correction formula now becomes:

$$A := r_{by} + r_{ba} \cdot r_{ya} (u^2 - 1)$$

$$B := \sqrt{[1 + r_{ab}^2 \cdot (u^2 - 1)] \cdot [1 + r_{ya}^2 \cdot (u^2 - 1)]}$$

$$r_c := \frac{A}{B}$$

where:  $r_{by}$  is the correlation between variables 2 and 3.  $r_{ba}$  is the correlation between tests 1 and 2.  $r_{ya}$  is the correlation between variables 3 and 1.  $u$  is the ratio of the unrestricted to restricted standard deviations in test 1.

With the correction formulas established there also came an inherent resistance to the process.

Researchers have not arrived at a general consensus about the desirability of adjusting correlations to obtain estimates for the unselected, rather than the selected population (Linn, Harnish, & Dunbar, 1981).

A number of researchers have questioned the accuracy of range correction when the assumptions are not met. Lord and Novick, for example, (1968) discovered a tendency for data to violate both linearity and homoscedasticity assumptions under severe selection. However, use of a Pearson correlation (corrected or uncorrected) is not recommended when these assumptions are violated.

Another concern is potential overcorrection. Novick and Thayer (1969) warned against this but provided no evidence of consistent overcorrection. In a like manner, Campbell (1976), discouraged the use of range correction without providing analytical or empirical evidence.

Seymour (1988), a lawyer, calls the correction for range restriction a “hydraulic function” that has the effect of increasing the size and significance of any correlation. Seymour further states that the Society for Industrial and Organizational Psychology (SIOP) has made an error stating that much of the differences in observed outcomes of validation research can be attributed to statistical artifacts. This statement comes after the American Psychological Association, Division 14, SIOP, had, in 1980, stated the correction of sample correlations for range restriction is a widely accepted practice in the psychological testing and measurement literature. Further, Seymour is unaware of the work of Levin (1972) that shows that corrections can and do decrease observed correlations.

Many researchers have observed the effects of correcting for range restriction when dealing with only a single variable, but few have discussed the effects of multiple variables upon range restriction, and none have discussed the accuracy of the multivariate correction. These effects motivate this paper.

The purpose of the current paper is to evaluate the estimation properties of correlations subjected to multivariate range restriction and multivariate correction. The average value of the corrected correlations will be compared to the known values and a regression model will be used to predict deviation of the corrected correlations from the population value. This will be accomplished by Monte Carlo simulation.

## Method

### Independent and Dependent Variables

There were four independent variables that were varied: sample size, magnitude of the specified correlation, number of variables, and mean selection ratio. The correlations were set equal to the broad range of values of: .10, .30, .50, .70, and .90. Sample size ( $n$ ) was varied to be 35, 140, 560, or 2,240 in accordance with the inverse square law. The number of variables was 5, 7, 9, and 11. There was curtailment upon 4, 6, 8, and 10, of the variables respectively. The variable not restricted was the criterion. The mean selection ratio (MSR) was set to .05, .15, .50, .85, or .95. The number of iterations, or samples, in each portion of the simulation was set at 35 based on the Theorem of Central Limits.

The first dependent variable was the difference between the uncorrected correlation and the specified correlation (rminruc). The second dependent variable was the difference between the corrected correlation and the specified correlation (rminrc).

#### Computer Program

Data were simulated by a program called "SIMLAW." This was developed by Douglas Jackson and Malcolm Ree (1990) and enhanced by Paul Rioux, Bill Myers, and Chris Lennon. SIMLAW is a computer software package comprised of Pascal programs designed to run on an IBM-compatible computer.

The simulation began with the specification of a mean selection ratio and the correlations among the variables. Next, four passes were made at the data while varying sample size (35, 140, 560, and 2,240).

With each pass of the data, several statistics were computed including: mean selection ratio, sample correlation uncorrected, sample correlation corrected, and the difference between the specified and the sample correlation before and after correction.

#### Results

The simulation results indicated that on average, the correction was downwardly biased and produced conservative estimates of the population values. The difference between the observed uncorrected correlation and the specified correlation was -.13 on average. When overcorrection occurred, the mean difference between the corrected correlation and the specified correlation was only -.004! See Table 1 for statistics. Appendix A presents histograms of the two dependent variables, rminrc and rminruc.

A linear model was then used to demonstrate and analyze what accounted for the average difference of -.13 between the uncorrected correlations and the specified correlations. The software used was SPSS which is a comprehensive and flexible statistical analysis and data management system.

The first regression performed included the independent variables of n, number of variables, MSR, and specified correlation. In the first set of regressions the dependent variable was the difference between the specified correlation and the uncorrected correlation. The respective Beta weights were -.028, -.001, -.671, and .375. The adjusted  $R^2$  was .58 showing the difference between the specified correlation and the uncorrected correlation depends heavily upon the variables MSR and the specified correlations. The adjusted  $R^2$  is Wherry's correction for the degrees of freedom that corrects for an upward bias in  $R^2$ .



The next regression included an interaction as an independent variable. This interaction was between MSR and the specified correlation. The regression resulted in the Beta weights of .750 for the specified correlation, -.667 for the interaction variable and -.190 for the MSR variable. The adjusted  $R^2$  was .66. The next step was to calculate the significance of the full and the reduced regressions by testing each of the linear models. In the full regression, which included the three variables MSR, the specified correlation, and the interaction variable, the  $R^2$  was .59054. In the reduced regression model which had only the two variables of MSR and the specified correlation, the  $R^2$  was .58848. An  $F$  test was calculated to compare the  $R^2$ s. The  $F$  test resulted in 59.341. Hence, the resulting linear regression was

$$Y = .03346 + -.060936 * (\text{MSR}) + -.306533 * (\text{Interaction of } n \text{ and the specified correlation}).$$

The next regressions were performed to predict the difference between the specified and the corrected correlations. The independent variables were MSR, the specified correlation,  $n$ , and the number of variables. This regression resulted in an  $R^2$  of .40757 with Beta weights of -.13 for MSR, -.21 for the specified correlation, -.31 for  $n$ , and .06 for the number of variables.

The next regression added interaction variables. The first interaction variable ( I1 ) was built by multiplying the mean selection ratio by the specified correlation. The second interaction variable ( I2 ) was created by multiplying the sample size by the specified correlation, and the third interaction ( I3 ) was constructed by multiplying  $n$  by MSR. Also constructed was a three way interaction ( I4 ) created by multiplying MSR, the specified correlation, and  $n$ .

The next regression included the dependent variable of the specified correlation minus the corrected correlation. The independent variables were MSR, specified correlation,  $n$ , number of variables, I1, I2, I3 and I4. The  $R^2$  was .18024. Table 2 shows the variables and their respective Beta weights.

Table 2	
Variable	Beta Weights
MSR	-.232651
Specified Correlation	-.341865
$n$	-.579694
I 1	.078194
I 2	.232648
I 3	.115749
I 4	-.013589
number of variables	.060097

The Beta weights show that sample size, specified correlation, mean selection ratio and I2 were the most influential in predicting the differences.

The next step was the removal of I4 to evaluate of the effective predictiveness of the three way interaction.

The  $R^2$  was lowered by a trivial .00001. Other regressions were run which included the removal of I1 resulting in a difference of .00085, and the removal of I2 resulting in a difference of .01108. The removal of interactions were done sequentially. When an interaction was removed it was left out of the next model. The small differences in  $R^2$ s demonstrate the most important interaction variable in predicting the difference in the corrected correlation was the second interaction (I2), sample size multiplied by the specified correlation.

Tests for significance involving the interactions were not reported because of the collinearity which existed among the interaction variables. Whenever there are strong correlations among independent variables, any particular variable may explain uniquely very little of the variation in the dependent variable, even though together the variables explain much. (Agresti & Finlay, 1984).

#### Conclusion.

The most important finding of this study is that on average the correction was a downwardly biased estimator. It produced conservative estimates of the population values. The average difference found between the specified correlations and the *uncorrected* correlations was much larger (-.13) than the average difference found between the specified correlations and the *corrected* correlations (-.004). Additionally, it should be noted that the standard deviations of these differences were .119 and .018 respectively. The standard deviations of the corrected correlations were 85% smaller than the uncorrected correlations. Not only is a better point estimate achieved but the variability of the estimate is reduced. The correction for range restriction provides a superior estimate and should be used.

This study also demonstrated what affects the accuracy of the correction. This information should be used to guide research practice. The four most influential independent variables, in order of importance, were sample size, the magnitude of the specified correlation, the interaction of the magnitude of the specified correlation and the mean selection ratio, and sample size.

The most influential variable, as reflected by the Beta weights, was sample size. As sample size was reduced the discrepancy between the corrected correlation and specified correlation was increased. Studies with low sample sizes should be appropriately evaluated as part of a meta-analysis. The importance of a large sample size was demonstrated by Ree and Earles (1993). This paper verified that as sample size increases, the standard error and the confidence intervals are reduced. Table 2 illustrates this effect.

Table 2 $\rho=0$			$\rho=.9$		
N	SE	1.96 (SE)	N	SE	1.96(SE)
35	0.171	0.335	35	0.032	0.062
140	0.084	0.164	140	0.016	0.031
560	0.042	0.082	560	0.008	0.015
2240	0.021	0.041	2240	0.004	0.007

The 1.96 (SE) is a 95% confidence interval. This table is consistent with the inverse square law of sampling error.

The second most important variable was the magnitude of the specified correlation. Correcting stronger underlying correlations will result in more accurate corrections. For example, under the condition of severe restriction there was a difference in correction accuracy of .03 when the specified correlations differed from .1 to .9. Table 2 demonstrates that as the correlation increases the standard error and the confidence interval are decreased. The more reliable the measurements, the more likely the researcher is to observe higher correlations.

The mean selection ratio is also a very influential variable. The more severe the MSR, the greater the difference between the uncorrected correlation and the specified correlation. When selection is severe and relatively few candidates are selected the correction is likely to be less accurate than when the MSR is low. Although the difference is not large and is not seen until the hundredths column, the difference between the corrected correlations exist. For example, when the MSR was set at .05, as compared to .95, the difference was seen to be .02. This discrepancy is much smaller than not correcting, -.13

Also important is the interaction of the magnitude of the specified correlation and sample size. A better estimate of the population will be found in a study that has a large sample size and high correlations. Conversely, a study with a small sample size and low correlations will result in less effective corrections

and poor estimates of the population parameter. In this case even though a less accurate correction could be expected, practitioners are better off using corrected correlations than uncorrected correlations.

The results also demonstrate the number of variables does not affect the accuracy of the correction. This means that the accuracy of the correction is independent of the number of variables on which there was a restriction and correction.

An ideal study would possess high correlations, low selection ratio, and a large sample size. This will result in more effective corrections and better estimates of the population parameter.

In summary, the correction for range restriction is a downwardly bias estimator that provides results that are more accurate than not correcting. This is important because range restriction occurs in numerous settings in applied psychology such as test validation, personnel selection, construct validation and job performance measurement.

Table 1

Descriptive statistics for the data.

Number of valid observations (listwise) = 400.00

Variable RMINRUC (Specified correlation minus the uncorrected specified correlation)

Mean	-.130	Std Dev	.119
Variance	.014	Kurtosis	6.441
S.E. Kurt	.243	Skewness	-1.557
S.E. Skew	.122	Range	1.047
Minimum	-1.0080	Maximum	.0390
Valid observations -	400	Missing observations -	0

Variable RMINRC (Specified correlation minus the corrected specified correlation)

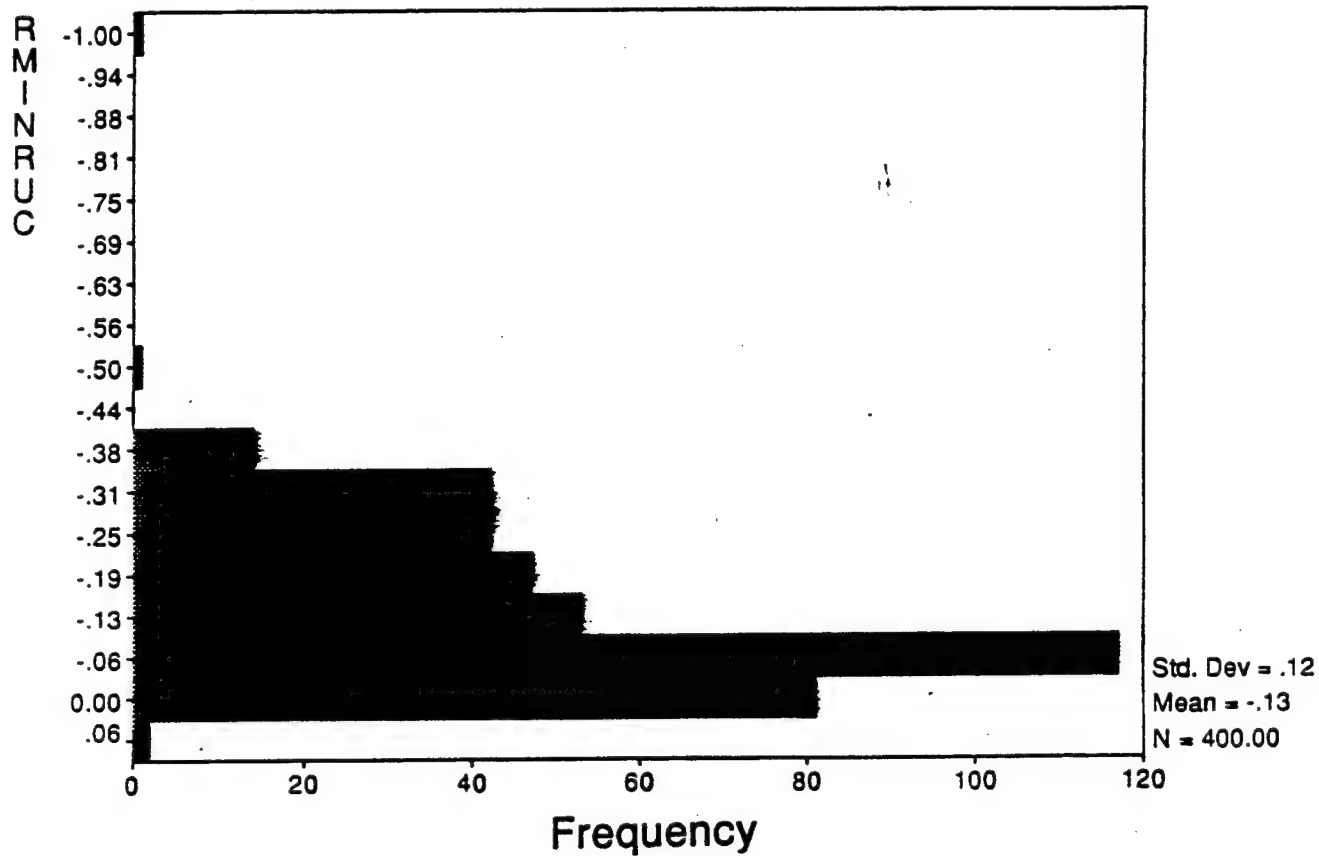
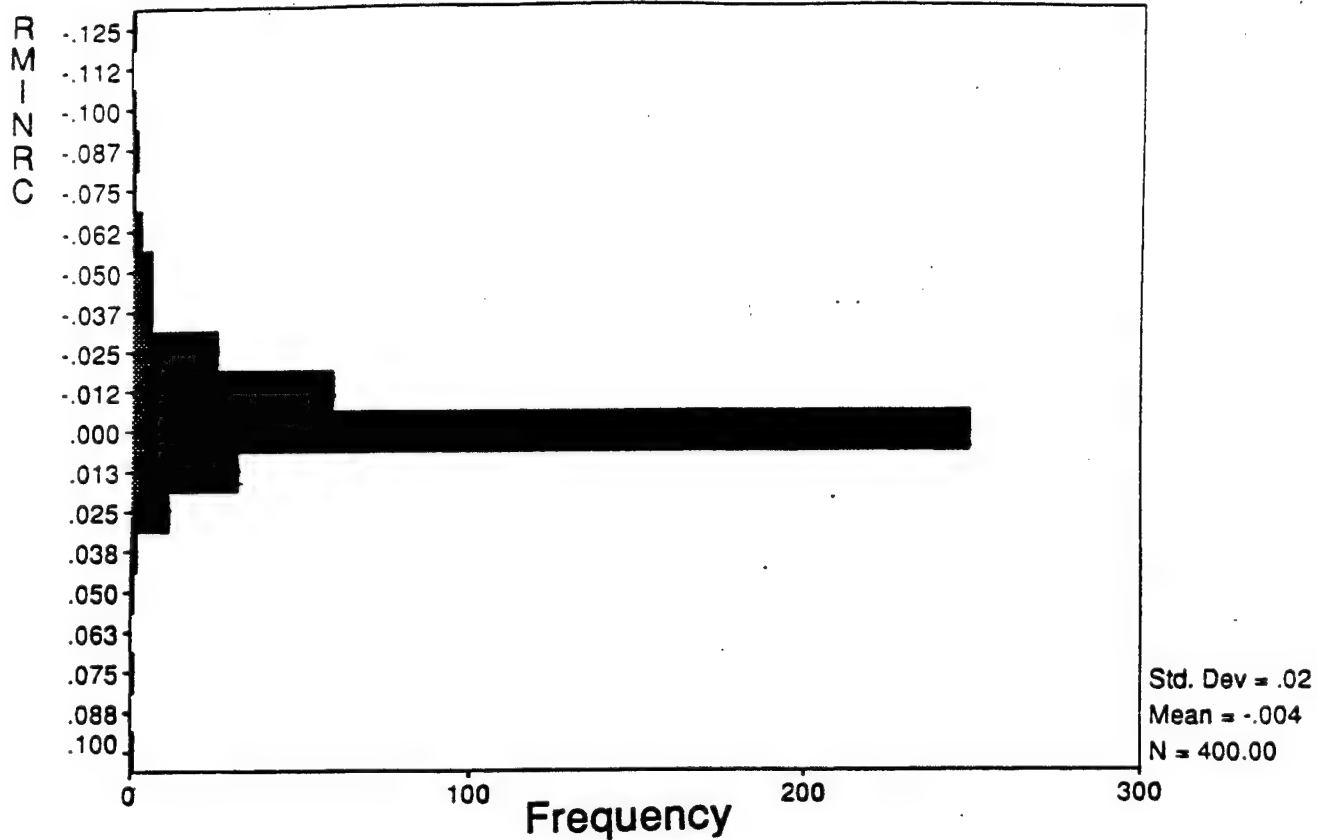
Mean	-.004	Std Dev	.018
Variance	.000	Kurtosis	11.581
S.E. Kurt	.243	Skewness	-1.448
S.E. Skew	.122	Range	.217
Minimum	-.1190	Maximum	.0980
Valid observations -	400	Missing observations -	0

Variable RUC (specified correlation uncorrected)

Mean	.373	Std Dev	.255
Variance	.065	Kurtosis	-.920
S.E. Kurt	.243	Skewness	.513
S.E. Skew	.122	Range	.892
Minimum	-.0080	Maximum	.8840
Valid observations -	400	Missing observations -	0

-----			
Variable	<u>RC</u>	(corrected specified correlation)	
Mean	.493	Std Dev	.285
Variance	.081	Kurtosis	-1.299
S.E. Kurt	.243	Skewness	.020
S.E. Skew	.122	Range	.992
Minimum	.0060	Maximum	.9980
Valid observations -	400	Missing observations -	0

# Appendix A



## References

- Agresti, L. J., & Finlay, M. H. (1986). Statistical methods for the social sciences (2nd ed.). California: Dellen Publishing Company.
- Campbell, J. P. (1976). Handbook of industrial and organizational psychology. Illinois: Rand-McNalley.
- Jackson, D. E., & Ree, M. J. (1990). Tool for studying the effects of range restriction in correlation coefficient estimation (AFHRL-TP-90-6). Brooks Air Force Base, TX: Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Research Division.
- Lawley, D. (1943). A note on Karl Pearson's selection formulae. Royal Society of Edinburgh: Proceedings (Section A: Mathematical and Physical Sciences), 62, 28-30.
- Levin, J. (1972). The occurrence of an increase in correlation by restriction of range. Psychometrika, 37, (1), 93-97.
- Linn, R. L., Harnish, D. L., & Dunbar, S. B. (1981). Corrections for Range Restriction: An empirical investigation of conditions resulting in conservative corrections. Journal of Applied Psychology, 66, 655-663.
- Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Mass.: Addison-Wesley.
- Novick, M. R., & Thayer, D. T. (1969). An investigation of the accuracy of the Pearson selection formulas. New Jersey: Educational Testing Service.
- Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. Journal of Applied Psychology, 79, 298-301.
- Ree, M. J., & Earles, J. A. (1993). g is to psychology what carbon is to chemistry: a reply to Sternberg and Wagner, McClelland, and Calfee. Current Directions in Psychological Science, 2, 8-9.
- Seymour, R. T. (1988). Why plaintiffs' counsel challenge tests, and how they can successfully challenge the theory of "validity generalization". Journal of Vocational Behavior, 33, 331-364.
- Thorndike, R. L. (1949). Personnel selection. New York: John Wiley & Sons.



**DISTRIBUTION-BASED EVALUATION AND ASSESSMENT  
OF MISSION READINESS FOR THE EVALUATION  
OF PERSONNEL TRAINING**

**David J. Woehr  
Associate Professor  
Department of Psychology  
Texas A&M University  
College Station, TX 77843-4235**

**and**

**Michael J. Miller  
Department of Psychology  
Texas A&M University  
College Station, TX 77843-4235**

**Final Report for:  
Summer Faculty Research Program  
Armstrong Laboratory**

**Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, Washington, D.C.**

**August 1995**

DISTRIBUTION-BASED EVALUATION AND ASSESSMENT  
OF MISSION READINESS FOR THE EVALUATION  
OF PERSONNEL TRAINING

David J. Woehr  
Associate Professor  
Department of Psychology  
and  
Michael J. Miller  
Department of Psychology  
  
Texas A&M University

Abstract

The present paper briefly summarizes a research project focusing on ways to improve the usefulness of organization level outcome measures of unit readiness/effectiveness for the evaluation of personnel training interventions. Toward this goal, a measurement approach using organization level outcome measures is presented. It is suggested that an adaptation of this approach has the potential to improve the utility of organization level criterion measures. In addition, potential sources of data for the evaluation of training interventions at the organization level are identified and evaluated.

# DISTRIBUTION-BASED EVALUATION AND ASSESSMENT OF MISSION READINESS FOR THE EVALUATION OF PERSONNEL TRAINING

David J. Woehr

and

Michael J. Miller

## Introduction

A vital concern for the Air Force is the maintenance of mission capability and readiness. A crucial mechanism for the maintenance of mission readiness is personnel training. There is little if any dispute that effective personnel training serves to enhance the effectiveness and capability of the Air Force in general. This fact is reflected in the overwhelming scope of training conducted throughout the Air Force and the tremendous amount of time and resources committed to the training endeavor.

Of tremendous importance to the design, implementation, and revision of training throughout the Air Force, as with any organization, is the ability to evaluate the effectiveness of training interventions. Goldstein (1991) defines training evaluation as: "the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to selection, adoption, value, and modification of various instructional activities." (p. 557) More specifically, the effective evaluation of any training intervention is crucial to informed decision making regarding the intervention. Central to effective training evaluation is the standard or criteria against which the training is evaluated. In addition, the comprehensive evaluation of training interventions mandates the use of multiple criterion measures. The impact of training interventions must be assessed at different levels (e.g., person, work group, organization). Unfortunately, organization level outcome measures are often dismissed as criterion measures due to contamination by extraneous aspects of the work environment. Despite this limitation, the use of these measures is extremely important for demonstrating the utility of training interventions.

### Organization Level Criterion Measures

Organization level outcome measures represent global indices of job effectiveness. They typically include results-oriented measures such as quality control indices, productivity or maintenance indices, promotion rate, salary progression or level, and turnover rates. The value of these measures as a standard for training evaluation is somewhat controversial. Two schools of thought can be found in the literature with respect to ways of conceptualizing the criterion construct. One school of thought emphasizes a conceptualization of performance as reflected in overt individual behaviors (e.g., Campbell, et al. 1970; Borman, 1983). This view focuses on the identification of behavioral regularities important to organizational functioning. The other school of thought focuses on outcomes. This view emphasizes the importance of outcomes and results to organizational functioning. Recent theories of the criterion construct, however, have begun to recognize the inextricable relationship between job behaviors and outcomes. Along these lines Binning and Barrett (1989) argue: "... optimal description of the performance domain for a given job requires careful and complete delineation of valued outcomes and the accompanying requisite behaviors" (p. 486).

### Problems with Outcome-Based Criterion Measures

The detailed conceptual delineation of the relationship between job performance and outcomes is especially relevant to training evaluation. An important direction for future research is a focus on behavior/outcome linkages and generating empirical support for these linkages. Unfortunately, the operationalization of specific outcome measures generates somewhat of a dilemma for training evaluation. On the one hand, the ultimate value of training lies in its ability to impact outcomes of value to the organization. Outcome measures (eg., productivity levels, turnover rates, error rates, etc.) at both individual and aggregate levels would appear to be the ultimate criterion of interest for evaluating training interventions. On the other hand, these measures suffer from a number of problems that limit their usefulness as a standard against which to judge the impact of training.

First and foremost among these problems is the fact that these measures are typically contaminated to an undetermined extent by sources of variance over which the individual has no control. Specifically, the measured outcome is to some extent determined by factors other than

individual performance. A second problem with outcome measures is that they are not based on a common metric. Outcome measures are often unique to particular units within an organization and thus are difficult to interpret and compare across organizational work groups or divisions. Additionally, the lack of a common metric typically precludes the meaningful aggregation of performance information across organizational units. A third problem is that these measures only provide an indication of outcome as opposed to the process underlying the outcome. Thus these measures provide little, if any, information about the nature of performance. Finally, the traditional use of outcome measures offers little, if any, means of assessing measurement quality (i.e., how good are the measurements obtained with these measures).

Thus, although regularly collected and typically readily available, as a criterion against which to judge the impact of various training interventions in organizations, outcome measures have not proven as useful as criteria which are defined in terms of individual behavior. Despite this, however, the use of these measures is extremely important for demonstrating the ultimate utility of training interventions. Consequently, an important goal with respect to training evaluation is the development of ways to improve the utility of organization level criterion measures. Toward this end, a specific measurement approach to outcome-based criterion measures is presented below.

#### A Distributional Approach to Criterion Measurement

The measurement approach presented here extends the system for assessing individual performance developed by Kane (1986) to outcome level criteria measurement. It is believed that this approach may offer a partial solution to the problems associated with outcome measures. The original system presented by Kane (1986), labeled Performance Distribution Assessment (PDA), is based on the distributional measurement model postulated by Kane and Lawler (1979). An important characteristic of this model is a focus on the range of performance observed. Specifically, the model stipulates that not only is the level of performance important, but the fluctuation or variance in performance must also be considered. For example, two individuals may both be appropriately characterized as "average performers"; however, if one is consistently average and the other alternates between very poor and very good, very different

pictures emerge with respect to the individuals' performance. Thus performance measurement must assess the range of performance over time. Specifically, performance is defined in terms of the outcomes of job functions that are carried out on multiple occasions within a specified time span (i.e., iterated job functions). It is expected that, due to varying levels of individual ability and motivation as well as varying levels of external constraints, these outcomes will reflect different levels of effectiveness. Performance can subsequently be represented in terms of the frequency at which various outcome levels occurred within a given time span.

Another important characteristic of the PDA approach is that it incorporates a relativistic scaling of performance information. More specifically, performance is expressed as a ratio of actual performance (as reflected in the performance distribution generated) to a maximum feasible performance distribution. This maximum feasible distribution reflects the highest level of performance attainable given the constraints under which the work occurs. This scaling process serves to express performance in terms of a relative range of potential performance. Thus, the method allows for quantifiably excluding from consideration in the evaluation of performance the range of performance that is attributable to circumstances beyond the performer's control.

The representation of performance in distributional form along with relativistic scaling has several important advantages. First, it allows for a consideration of performance variability as well as average levels of performance. Thus it allows for an assessment of the consistency of performance and the extent to which negatively valued outcomes are avoided. In this way more information is provided regarding the idiosyncratic nature of individual performance. Second the relativistic scaling process advocated by the PDA process produces measures of the effectiveness of performance on relativized 0-100% scales with common zero and common upper limits of 100%. Thus any given percentage level remains constant in its meaning regardless of the job, division, or even the organizational level in which it occurs. At the same time, the particular outcome measures used to assess performance may be individualized to meet situational demands and organizational constraints. Specifically, if positions have appreciably different content and extraneous-constraint conditions, measures can be scaled to account for these differences.

The PDA approach was originally advocated as method for enhancing performance ratings. Specifically, it was formulated to incorporate subjective estimates of individual performance outcome frequencies (i.e., supervisory ratings of the frequency at which individuals performed at a particular level). However, its focus on the frequency of particular performance outcomes make it particularly amenable to use with more objective outcome measures. Thus, the application of this methodology to the measurement of organizational outcomes using iterative operational measures appears to be a fruitful avenue for research and may serve to increase the utility of these measures in the training evaluation process.

#### Adaptation of the PDA Approach for Outcome Level Measures

As noted above, the PDA system appears to be well suited for the measurement and scaling of operational criterion measures. For purposes of illustration, Table 1 presents hypothetical evaluation data presented in PDA format. In Table 1 the outcome range represents 5 equidistant steps between the highest possible performance outcome (listed as 95 in the Table) and the lowest acceptable performance level (listed as 75 in the Table). These values are based on supervisory or SME estimates. The utility weights represent the utility or value to the organization of performance at each of the 5 levels. Once again these values are based on SME estimates. Supervisors/SMEs also provide estimates for the "maximum" feasible distribution of performance as well as estimates of the "actual" performance level distribution of the individual being evaluated. Other values in the Table are calculated from the supervisor estimates. Here, it should again be noted that the PDA process was originally intended as a subjective performance rating system. Thus many of the data distribution utilized are based on supervisor or SME estimates of both subordinate performance and minimum and maximum performance levels. An adaptation of PDA for use with operational measures does away with these subjective estimates and replaces them with actual frequencies based on archival records of the operational measures. Also, the maximum feasible distribution is replaced with a "benchmark" distribution. This "benchmark" distribution may represent either an estimated ideal distribution of performance or the actual performance distribution of a comparison unit (i.e., an earlier time frame or another work unit). This revised approach is labeled here as Distribution-Based Evaluation and Assessment of Mission Readiness (DEAMR). This approach extends the

beneficial characteristics of relative distribution based performance assessment to organization level outcome measures. More specifically, characteristics of the DEAMR process include:

1. Performance measurement is relativistic. Outcome measures are scaled relative to maximum possible and minimum acceptable performance levels. Performance distributions are relative to some "benchmark" distribution. Thus, measurement considers the extraneous factors that may influence outcome measures.
2. Performance measurement is based on common metric. All measures are expressed in terms of percentages and thus have minimum and maximum points.
3. Multiple measures of performance are provided; performance is described in terms of mean level, consistency, and negative range avoidance. These multiple measures provide more information about the nature of performance and performance problems.

Another important characteristic of the DEAMR system is that it is easily automated. Relatively little data is required in order to calculate the distributional parameters. This data required includes the highest possible and lowest acceptable performance level, an estimate of the utility weight associated with the lowest acceptable performance level, and the actual frequency of performance outcomes at each of the performance levels. Table 2 represents the output of a spreadsheet based program specifically designed to provide distributional performance information. The highlighted boxes indicate where data must be input into the program. Performance distribution information is then automatically calculated and displayed both numerically and graphically.

#### Illustration of the DEAMR Process with Actual Data

An important aspect of the present study is the identification of potential data sources as well as the development of a system for improving the utility of outcome measures. Toward this end several existing Air Force data sources were identified and examined within the DEAMR framework.



### CAMS Data

An initial potential data source examined was based on jet aircraft maintenance data gathered through the Core Automated Maintenance System (CAMS). This data provides the advantages of being regularly and systematically collected as well as being readily available in a usable format. The data examined in the present study was an archival six month sample (April - September 1990) of F15 maintenance data. CAMS data represents a record of aircraft maintenance activities performed. Table 2 presents an example of the output of this program following the DEAMR methodology. The primary CAMS based measure used in the DEAMR analysis is time to completion on various job tasks. Time to completion is based on the difference between the start and stop times for the task. Table 3 presents the DEAMR output for the "Benchmark and Repair" action taken category derived from the CAMS data. Time to completion for job functions in this category ranged from .5 to 8. Distributional characteristics of performance in this category are presented in Table 3.

### Limitations of CAMS Data

DEAMR analysis of the CAMS based time to completion data revealed a number of potential limitations of this data source. Some of these limitations were specific to the sample of data available. These included:

1. The data set was not collected or developed for this application (i.e., the 6 month sample had been culled from the larger data source for another purpose).
2. The number of cases available for analysis was quite small, thus making it difficult to make generalizations.
3. The units of comparison used were fairly broad (i.e., performing work center) making determination of appropriate benchmark distributions difficult.
4. The action category breakdowns are overly broad in that they do not consider on which equipment or system the action is performed.

Other limitations of the CAMS data resulted more generally from the nature of the data itself. These limitations represent a more serious threat to the usefulness of the CAMS based measures and include:

1. The CAMS-based measure is based purely on maintenance data and thus does not take into account unit characteristics such as manpower, mission requirements, etc.
2. The CAMS data is contaminated to an unknown extent by several factors including: the individual performing the work may not enter the data; standard or benchmarked time estimates may be entered instead of actual times; and, the measures are very heavily equipment driven.

### Current Data Options

Given the limitations identified with the CAMS data, work was undertaken to locate other, more up to date, potential data sources. Two other current data sources were identified.

### Mission Capability Data

A review of operational measures routinely collected at the 56th fighter wing at Luke AFB, Az. revealed one potentially valuable data source. This data represents a combination of CAMS- based maintenance data and unit mission characteristic data. This combination of data overcomes the major limitation associated with using CAMS data alone. More specifically, both equipment and unit mission and manpower characteristics are considered. Example measures include fully mission capable rate (FMC), man hours per flying hours, air and ground abort rates, etc. A listing of the actual data collected is presented in Table 4. The data presented in Table 4 is used to compute direct indices of mission capability/readiness (presented in Table 5). The indices presented in Table 5 have many desirable characteristics with respect to use with the DEAMR system. These characteristics include:

1. The measures are regularly and systematically collected.
2. It appears that these indices are both required by and reported to MAJCOM. Thus it is likely that these measures are available Air Force wide.
3. The mission capable/readiness indices reflect both equipment, mission, and manpower characteristics.
4. The indices are easily aggregated from the individual unit level to higher levels of the organization (wing, command, etc.).

5. The indices reflect multiple measures of performance within a specified time span (iterated job function) and thus are readily amenable to the DEAMR system.

These characteristics suggests further examination of these indices as criterion measures for training evaluation is warranted.

#### Quality Assurance Program Data

Another potentially valuable data set for use with the DEAMR system is quality assurance program data. Initial inquiry revealed that the 56th fighter wing has a wing based quality support office. This office regularly collects and disseminates quality control data relevant to fighter aircraft maintenance. This data is largely based on results of scheduled quality assurance inspections. Two indices appear to be particularly relevant for training evaluation. The first is quality assurance inspection pass rates. The second is the number of discrepancies per inspection. Discrepancies represent negative findings during quality assurance inspections. Both inspection pass rates and discrepancies are grouped by TEC code indicating the aircraft system or maintenance activity to which they apply. This data already provides valuable feedback to the training function and thus further examination of these measures as criterion measures for training evaluation is warranted.

#### Using DEAMR for Training Evaluation

Both the mission capability/readiness indices and quality assurance data represent viable potential criterion measures for the evaluation of training effectiveness at the outcome level. Further both data sources meet the requirements for use with the DEAMR system. Thus it is possible to rescale the data in distributional form. The DEAMR format could then be used to evaluate the effectiveness of specific training interventions.

While the data sources identified generally provide a potentially valuable source of information with respect to training evaluation, it is important to consider the further refinement of the data. More specifically, it would be beneficial to establish a pool of potential indices most relevant to specific training interventions to be evaluated (eg., maintenance indices for maintenance technician training). Here it is important to identify and evaluate key measures

from the larger pool of potential measures. The focus of this measure evaluation would be to identify indices that are: a) important to unit effectiveness, b) frequently and reliably collected, c) sensitive to individual performance, d) relatively insensitive to system variables, and e) relevant to training intervention. Information about outcome indices may be obtained through either SME workshops or through structured questionnaires. SME's would be used to provide information about each potential measure (eg., the relative importance of each measure, sensitivity to individual performance) as well as information relevant to the DEAMR process (eg., item utility weights, optimal possible and minimal acceptable levels, etc.). It is only through such systematic examination of the measures available that detailed conceptualizations of the linkages between individual performance and organizational outcome can be established. Ultimately, the effective use of outcome measures for the evaluation of personnel training depends on the delineation of these linkages.

## References

- Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74, 478-494.
- Campbell, J.P., Dunnette, M.D., Lawler, E.E., & Weick, K.E. (1970). Managerial Behavior, Performance, and Effectiveness. New York: McGraw-Hill.
- Goldstein, I.L. (1991). Training in work organizations. In M.D. Dunnette & L.M. Hough (Eds.) Handbook of Industrial and Organizational Psychology, vol. 2, (pp. 507-620). Palo Alto: Consulting Psychologists Press.
- Borman, W.C. (1983). Implications of personality theory and research for the rating of work performance in organizations. In F. Landy, S. Zedeck, & J. Cleveland (Eds.) Performance measurement and theory (pp. 127-172). Hillsdale, NJ: Erlbaum.
- Kane, J.S. (1986). Performance distribution assessment. In R. Berk (Ed.), Performance Assessment: Methods and Applications, (pp. 237-273). Baltimore, MD: Johns Hopkins University Press.
- Kane, J.S., & Lawler, E.E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B. Staw (Ed.) Research in Organizational Behavior, Vol 1. Greenwich, CT.: JAI Press.

**Table 1: Hypothetical Performance Distribution Data**

Levels	1	2	3	4	5		Mean (SD)	Negative Range Score
Outcome Range	75	80	85	90	95		----	----
Utility Wts	-150	-87.5	-25	37.5	100		----	----
Maximum	0	.10	.15	.15	.60		53.13 (65.18)	-12.50
Actual	0	.10	.20	.30	.40		37.50 (62.50)	-13.75
Most Consistent	0	0	.20	.60	.20		37.50 (39.53)	- 5.00
Least Consistent	.067	.066	.267	0	.60		37.50 (82.30)	-22.50

EFFECTIVENESS OF THE ACTUAL DISTRIBUTION IS ESTABLISHED RELATIVE TO THE MAXIMUM FEASIBLE DISTRIBUTION - E.G.:

Relative Effectiveness of Mean of

Actual Distribution = (37.50/53.13) = 71%

Relative Effectiveness

of Consistency of (82.30 - 62.50) /

Actual Distribution = (82.30 - 39.53) = 46%

Relative Effectiveness

of Negative Range

Avoidance of Actual (-22.5 - (-13.75)) /

Distribution = (-22.5 - (-12.50)) = 88%

Distribution Characteristics Calculations for:  
Unit: OVERALL  
Measure: Calibrate

Perf. Level	Perf. Range	Utility Weights	Performance Level Frequency	Performance Level Percentage	Distribution Characteristics	Utility Wt. Scale	PL Scale
1	8.00	-100.00	6.00	0.12	Mean =	58.00	4.16
2	6.13	-50.00	1.00	0.02	SD =	66.60	1.33
3	4.25	0.00	2.00	0.04	Skewness =	-1.56	-1.56
4	2.38	50.00	11.00	0.22	Kurtosis =	1.06	1.06
5	0.50	100.00	30.00	0.60	Negative Range Score =	-13.00	
Total Obs =			50.00	1.00			

**Step 1**  
Enter min and max performance levels

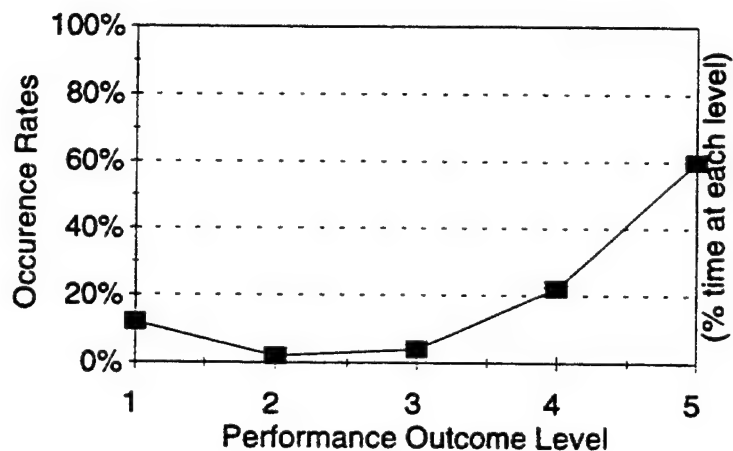
**Step 2**  
Enter min. and max utility weights

**Step 3**  
Enter freqs. for each perf. level

Perf level % are calculated

Distributional Characteristics are calculated

### Performance Outcome Distribution



Distribution Characteristics Calculations for:  
Unit: OVERALL  
Measure: Bench Check  
and repair

Perf. Level	Perf. Range	Utility Weights	Performance Level Frequency	Performance Level Percentage	Distribution Characteristics	Utility Wt. Scale	PL Scale
1	8.00	-100.00	3.00	0.15	Mean =	30.00	3.60
2	6.13	-50.00	1.00	0.05	SD =	65.95	1.32
3	4.25	0.00	2.00	0.10	Skewness =	-0.94	-0.94
4	2.38	50.00	9.00	0.45	Kurtosis =	-0.30	-0.30
5	0.50	100.00	5.00	0.25	Negative Range Score =	-17.50	
Total Obs =			20.00	1.00			

**Step 1**  
Enter min  
and max  
performance  
levels

**Step 2**  
Enter min.  
and max  
utility weights

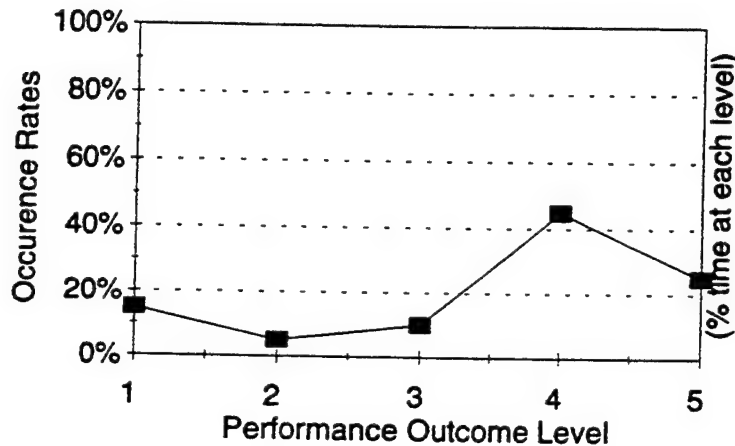
**Step 3**  
Enter freqs.  
for each perf.  
level

Perf level %  
are calculated

Distributional  
are

Characteristics  
calculated

### Performance Outcome Distribution





**Table 4: Mission Capability Measures Collected**

<b>Data</b>	<b>Source</b>
Authorized Aircraft	Wing Plans & Scheduling
Possessed Aircraft (aircraft available to flight, w possession codes of CA, CB,TF, or ZB0)	Wing Plans & Scheduling
Total Aircraft Possessed Hours (APH)	
Total Sorties Flown	
Total Man Hours	
Total Flying Hours	
Fully Mission Capable Hours (FMC)	CAMS # 459
Non Mission Capable Supply Hours (NMCS)	CAMS #459
" " " Maintenance (NMCM)	CAMS #459
" " " Both (NMCB)	CAMS #459
Partially Mission Capable Supply Hours (PMCS)	CAMS #459
" " " Maintenance (PMCM)	CAMS #459
# of 4/8/12 Hour Fixes	CAMS #460
# of Code 3 Breaks	CAMS #174
# of Ground Aborts	CAMS #174
# of Air Aborts	CAMS #174
# of Repeats (same malfunction on next flight)	CAMS #174 & FS Recap Sheets
# of Recurs (same malfunction w/in next 3 flights)	CAMS #174 & FS Recap Sheets
# Aircraft Awaiting Parts (AWP)	
# Aircraft Awaiting Maintenance (AWM)	
Total Points Earned	
Total Points Scheduled	
Local Sorties Scheduled	
Weather Adds/Deletes	
Ferry/FCF Adds/Deletes	
Other Adds/Deletes	

**Table 5: Mission Capability/Readiness Indices Calculated**

<u>Measure:</u>	<u>Formula:</u>	<u>Calculated by:</u>
Avg. Possessed Aircraft (APA)	$\frac{\text{Total Acft Possessed Hrs}}{\text{Total Days in Month} \times 24}$	FS Analysts (daily)
Actual UTE Rate	$\frac{\text{Total Sorties Flown}}{\text{APA Aircraft}}$	FS Analysts (daily)
Manhours per Flying Hours	$\frac{\text{Manhours}}{\text{Total Flying Hours}}$	
Adjusted Sortie Schedule (Adj. Sortie Sched.)	$\frac{\text{Local Sorties Scheduled} + \text{Weather Adds} + \text{Ferry/FCF Adds} + \text{Other Adds} - \text{Weather Deletes} - \text{Sympathy Deletes} - \text{Other Deletes}}{\text{Adj. Sortie Sched.}}$	
Aircraft Scheduling Effectiveness	$\frac{\text{Adj. Sortie Sched.} - \text{Chargeable Deviations}}{\text{Adj. Sortie Sched.}} \times 100$	
Maintenance Plan Rate	$\frac{\text{Total Pts. Earned}}{\text{Total Pts. Scheduled}} \times 100$	
Mission Capable (MC) Rate	$\frac{\text{PMCM} + \text{PMCS} + \text{PMCB} + \text{FMC}}{\text{APH}}$	
Fully Mission Capable (FMC) Rate	$\frac{\text{FMC}}{\text{APH}} \times 100$	
Non Mission Capable (NMC) Rate	$\frac{\text{NMCS} + \text{NMCM} + \text{NMCB}}{\text{APH}} \times 100$	
Total Non Mission Capable Maintenance (TNMCM) Rate	$\frac{\text{NMCM} + \text{NMCB}}{\text{APH}} \times 100$	
Total Non Mission Capable Supply (TNMCM) Rate	$\frac{\text{NMCS} + \text{NMCB}}{\text{APH}} \times 100$	
Non Mission Capable Both (NMCB) Rate	$\frac{\text{NMCB}}{\text{APH}} \times 100$	

**Table 5: Mission Capability/Readiness Indices Calculated (cont.)**

Measure:	Formula:	Calculated by:
Partially Mission Capable Both (PMCB) Rate	$\frac{\text{PMCB}}{\text{APH}} \times 100$	
Total Partially Mission Capable Maintenance (TPMCM) Rate	$\frac{\text{PMCB} + \text{PMCM}}{\text{PAH}} \times 100$	
Total Partially Mission Capable Supply (TPMCS) Rate	$\frac{\text{PMCB} + \text{PMCM}}{\text{A PH}} \times 100$	
Ground Abort Rate	$\frac{\text{Ground Aborts}}{(\text{LCL Sorties Fln} + \text{Grnd Aborts})} \times 100$	
Air Abort Rate	$\frac{\text{Air Aborts}}{(\text{LCL Sorties Fln} + \text{Air Aborts})} \times 100$	
Total Abort Rate	Grnd Abort Rate + Air Abort Rate	
Code 3 Break Rate	$\frac{\# \text{ Code 3 Breaks}}{\text{Total Sorties Flown}} \times 100$	
4/8/12 Hour Fix Rates	$\frac{\# \text{ of 4/8/12 Hour Fixes}}{\# \text{ Code 3 Breaks}} \times 100$	
Deferred Discrepancy Rate (repair which cannot be done w/in 5 days)	$\frac{\# \text{ of AWM/AWP}}{\# \text{ of possed aircraft}} \times 100$	
Repeat/ Recur Rate	$\frac{\# \text{ Repeats or } \# \text{ Recurs}}{\text{Local Sorties Flown}} \times 100$	

**TEST EQUITY OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB)  
CLASSIFICATION COMPOSITES**

**Krystin E. Mitchell  
Research Associate  
Department of Psychology**

**University at Albany  
State University of New York  
Albany, NY 12222**

**Final Report for:  
Graduate Student Research Program  
Armstrong Laboratory**

**Sponsored by:  
Air Force Office of Scientific Research  
Bolling Air Force Base, DC**

**and**

**Armstrong Laboratory**

**July 1995**

**TEST EQUITY OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB)  
CLASSIFICATION COMPOSITES**

**Krystin E. Mitchell  
Graduate Research Assistant  
Department of Psychology  
University at Albany**

**Abstract**

This study used the Cleary model (1968) to examine test bias in the Armed Services Vocational Aptitude Battery (ASVAB). The focus of this study was to assess test equity for white versus black airmen in four self-paced technical training courses. Measured predictor variables included the airmens' Mechanical (M), Administrative (A), General (G), and Electronic (E) aptitude composites derived from ASVAB subtest scores. The criteria for the present study were final course grade (FSG) and time to complete the training course (TTC). Results were generally consistent with past research on measures of general cognitive ability indicating that level bias (or a lack of bias) was more prevalent than slope bias (Sackett & Wilk, 1994). Additionally, the use of a common regression line when level bias was present led to the overprediction of criterion scores for black airmen.

## TEST EQUITY OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB) CLASSIFICATION COMPOSITES

Krystin E. Mitchell

The United States military uses the Armed Services Vocational Aptitude Battery (ASVAB) for applicant selection for enlistment and classification of enlistees into specific technical training schools. The use of cognitive ability tests for these purposes is quite common in the military and civilian sectors, although the use of these tests remains controversial (Humphreys, 1993; Jensen, 1980). The controversy concerning the use of cognitive ability tests stems from the consistent finding that blacks score an average of 1.2 standard deviations below whites on these measures (Jensen, 1980). However, mean differences on cognitive ability measures are not sufficient to indicate that a predictor is biased. Test bias is generally assessed using the Cleary model (1968). Cleary (1968) indicated that a test is biased if, 1) the intercepts between the majority and minority subgroups are significantly different and/or, 2) the slopes for the majority and minority subgroups are significantly different. Generally, there is little evidence that the slopes for the majority and the minority subgroups differ significantly on cognitive ability tests (Hunter, Schmidt, & Rauschenberger, 1975; Rothstein & McDaniel, 1991). When level bias is detected, the intercept for the majority group is generally higher than the intercept for the minority group (Sackett & Wilk, 1994). In this instance, a common regression line is used to predict all applicants' criterion scores. The use of a common regression line will overpredict minority applicants' scores, leading to a larger number of minorities being hired than if separate regression lines for each subgroup were used. Consequently, the use of cognitive ability tests in this case is not controversial because criterion scores for the minority group will be over, not under, predicted (Sackett & Wilk, 1994).

Although much research has addressed the issue of test equity as defined by the Cleary model, additional research is always valuable when assessing specific predictors and criteria of interest. Research on the ASVAB generally supports the overall validity of this measure for predicting final school grade (FSG), time to complete training (TTC) measured in days, and whether or not an individual will pass or fail training (for a review, see Welsh, Kucinkas, & Curran, 1990). Individual studies supporting the validity and equity of the ASVAB for blacks and whites include work done by Gordon (1953), Maier and Truss, (1984), McLaughlin, Rossmeissl, Wise, Brandt, and Wang (1984), and Wilbourn, Valentine, and Ree (1984). However, Welsh, Kucinkas, and Curran (1990) state that "...there is enough, though spotty evidence, of some under and overprediction of minority groups for some specific military occupations to warrant some more thorough investigations of the equity of Service-specific selector composites." (p., 106) (e.g., Wise, 1992)

The present study used the Cleary model (1968) to assess whether or not there was level bias or slope bias for the four Air Force composites of Mechanical aptitude (M), Administrative aptitude (A), General aptitude (G), and Electronic aptitude (E). The criteria used were FSG and TTC. The current study extends past research on the equity of the ASVAB by using the TTC criterion. The training courses in this study were "self-paced" courses where trainees could advance through training as quickly or slowly as they desired. The underlying assumption

when using aptitude to predict TTC is that in a self-paced environment, high ability trainees will progress through training more rapidly than low ability trainees. Thus, TTC is assumed to be under the trainees' direct control, whereas FSG may be more susceptible to criterion contamination due to rater bias. For example, the instructors of the training courses may give higher final grades to whites than blacks regardless of actual performance levels. Although the research on rater bias is mixed (Bretz, Milkovich, & Read, 1992), there is empirical evidence which suggests that the race of the ratee may influence performance ratings (Kraiger & Ford, 1985). A common complaint leveled against equity studies in general is that the criterion is assumed to be uncontaminated, although this is generally never assessed (Nunnally & Bernstein, 1994). The present study allows for the comparison of FSG and TTC in an equity study framework to assess whether rater bias will artificially increase the intercept differences between blacks and whites for FSG.

The hypotheses for the present study were threefold. First, it was predicted that level bias would be detected revealing that whites would have consistently higher predicted FSG scores across all levels of the relevant aptitude predictor. Additionally, it was hypothesized that use of a common regression line in this instance would overpredict the criterion scores for black trainees.

Second, it was predicted that there would be level bias for TTC. This level bias was predicted to indicate that blacks would require more days to complete a training course across all levels of aptitude. Third, the relative magnitude of level bias was predicted to be less for TTC than for FSG. This hypothesis was based on the assumption that rater bias in FSG would inflate the relative difference between racial subgroups on the criterion scores. Thus, the differences between the intercepts for blacks and whites on TTC were predicted to be smaller than the differences for FSG due to the lack of rater bias in TTC.

#### Method

##### Sample

The sample consisted of 26,285 Air Force enlistees who entered specific self-paced technical training schools between 1970 and 1983.

##### Procedure

Data for this study were extracted from archival files maintained by the Air Force.

Data file preparation. The archival files contained technical training course information from training courses dating from 1966 - 1993. Self-paced courses were identified through historical records. When historical records were not available, additional analyses were performed to identify self-paced courses. First, all outliers in the number of training days were identified and removed from the data file. A value was considered an outlier if the score represented a 25% increase or decrease from the nearest score in the distribution. Second, the coefficient of variation for the number of training days (standard deviation of the number of training days divided by the mean number of training days for each training course) was used to identify the degree of variability in the number of training days for a particular course. Courses with a coefficient greater than .15 were identified as being self-paced. Finally, all cases were removed from all of the remaining training courses which contained values of less than 50 percent for the final school grade in a particular course.

Training courses in the present study. As previously mentioned, all of the training courses used in to present study were selected because they were self-paced. Two additional criteria had to be met before a course was considered appropriate for the present study. First, the proportion of minority group members to majority group members was evaluated for each course. Courses were selected where the proportion of minority group members to majority group members was approximately 1:10 or higher to allow adequate power for analyses involving subgroup comparisons (Stone-Romero, Alliger, & Aguinis, 1994). Although a proportion of 1:2 would be ideal to ensure adequate statistical power, only one course would have met this criterion. Thus, the criterion of a proportion of 1:10 was adopted to ensure all the classification composites (MAGE) were represented. Therefore, four courses were selected to represent each of the MAGE predictor composites. The following courses met the criteria for inclusion in present study: Aerospace Ground Equipment training (1976-1979), Staff Support Administration training (1977-1983), Air Traffic Control Training (1971-1978), and Aircraft Electrical Systems Training (1970-1979).

#### Measures

Predictor variables. The predictor variables in the present study were the MAGE composites derived from the 10 subtests on the ASVAB. Prior to 1976, the Air Force used the Airman Qualifying Examination (AQE) to select and classify airmen into technical training courses. In 1976 the ASVAB replaced the AQE and since 1976 has been used to select and classify individuals across all four services. However, the MAGE composites derived from different cognitive ability tests have been shown to be highly correlated (Vitola & Alley, 1968). Moreover, the content of the ASVAB was revised during the time period covered by the present study. However, the MAGE composite scores were calibrated across forms and thus are assumed to be comparable in meaning for the time period covered in the present study.

Only the specific MAGE composite scores used to place recruits in particular technical training schools were used. Consequently, M scores were only used to predict Aerospace Ground Equipment trainees' performance, A scores were used to predict the criteria for Administrative support trainees, G scores were used to predict Air Traffic Control trainees' performance, and E scores were used to predict the performance of Aircraft Electrical System trainees.

Each composite was also associated with a different minimum cutoff score necessary for acceptance into a particular technical training course. The minimum required percentage score on the MAGE composites used during the time period for the self-paced courses offered were as follows: M = 40th percentile, A = 40th percentile, G = 60th percentile, E = 50th percentile. To assess the relationship between the predictor scores and the criteria across the range of interest for the aptitude composites, all scores below these minimum cutoff scores were deleted from the data file prior to the analysis of the data.

Performance criteria. The two criteria used in this study were final school grade (FSG) and time to complete training (TTC). FSG was measured as the percent correct for each trainee in a particular self-paced training course. FSG scores ranged from 61-99.



TTC was measured in days. TTC ranged from 15-751 depending on the particular training course evaluated.

#### Data Analysis

The hypotheses in the present study were tested using the General Linear Model (GLM) approach. This approach compares a "full" model which contains all the predictor variables of interest to "restricted" models which contain a subset of the variables in the full model (Pedhazur, 1982). In the present study, a series of four linear models were tested for each training course and each criterion. Model 1 represented the starting or "full" linear model. This model included the following predictor variables: 1) predictor scores for all trainees, 2) racial subgroup membership represented by binary coding, i.e., 0=black, 1=white and 3) the interaction term for variables one and two. The number of coded vectors necessary to represent racial subgroup membership is  $k-1$  where  $k$  is the number of levels of the potential moderator variable (Pedhazur, 1982). Thus, variable two for the present study was represented by one coded vector. Consequently, the interaction term, which is derived from multiplying the vector for variable one by the vector for variable two, contained predictor scores for white airmen only. The use of an additional coding vector for either variable two or the interaction term would only have provided redundant information (Pedhazur, 1982). The specification for model 1 allowed the regression slopes and intercepts to differ between the majority and minority subgroups.

Model 2 contained: 1) the predictor scores for all trainees and 2) racial subgroup membership. This model specified that the regression lines for the racial subgroups had the same slope, but the intercepts for the two subgroups were allowed to differ. Model 3 involved regressing FSG or TTC on one of the MAGE composites. Model 3 specified that the slopes and intercepts for the racial subgroups were equivalent. Model 4 was a null model which represented the null hypothesis that  $Rho = 0$ .

GLM first compares Model 1 (the full model) to the first restricted model (Model 2) to assess if there is a loss in predictive efficiency. The  $F$ -statistic is used to test the change in  $R$ -square value between these two models (Ward & Jennings, 1973). If the  $R$ -square value is found to be significant when Model 1 and Model 2 are compared, then slope bias is indicated. If the  $F$  value for this analysis is not significant, then Model 2 is compared to Model 3. If the change in  $R$ -square is found to be significantly different between Models 2 and 3, this indicates intercept differences between the racial subgroups. If this analysis is not significant, however, then Model 3 is tested for statistical significance against a null model positing  $Rho = 0$ . If Model 3 is found to be significantly different from the null model and no slope or intercept differences were previously found, then the test is valid predictor of the criterion regardless of subgroup membership.

#### Results

Means and standard deviations for each of the aptitude composites and FSG and TTC are shown in Table

1. The general pattern of the means between blacks and whites for the MAGE composite scores indicates

---

Insert Table 1 about here

---

that on average blacks had lower means (.36 s.d. to .80 s.d. lower) and lower standard deviations on the classification composites. Differences in FSG were less pronounced, with blacks having comparable standard deviations to whites and scoring only 1.16 (.20 s.d.) to 2.93 (.50 s.d.) grade points lower than whites. Means on the TTC criterion revealed that blacks required a comparable number of training days to approximately 8 more days (.66 s.d.) than whites to complete the training courses. The standard deviations for the racial subgroups for TTC were comparable.

Table 2 presents the corrected and uncorrected Pearson-Product Moment correlation coefficients between the MAGE composite scores and FSG and TTC for blacks and whites.

---

Insert Table 2 about here

---

The results show that the MAGE composite scores are significantly correlated with FSG and TTC in all cases except where A was used to predict TTC for white airmen. Additionally, the correlation coefficients for TTC were smaller in magnitude than the correlation coefficients for FSG for both racial subgroups.

Additionally, the relative magnitudes and significance levels of the correlation coefficients were generally larger for whites than for blacks across both criteria. However, the correlation coefficients for the subgroups may have differed because the black subgroups had substantially fewer people, thus limiting power and consequently the size of the correlation coefficients for the minority subgroup. All of the correlations in the present study were corrected for range restriction on the predictor variable. On average, this range restriction correction increased the magnitude of each correlation coefficient by approximately .10.

Table 3 shows the standard errors of the estimate (SEE) for both subgroups, and indicates the F statistic value which tests for the statistical difference between these estimates. For each of the training courses the SEE for

---

Insert Table 3 about here

---

the white subgroup was not significantly different from the SEE for the black subgroup. This indicates that the homogeneity of variance assumption was met, which is one of the primary statistical assumptions which should be met before for performing GLM (Pedhazur, 1982).

The results of the GLM analyses for each training course are presented in Table 4. This table shows the R

---

Insert Table 4 about here

---

squared values for each of the models that was tested and the resulting  $F$  values for the model comparisons. The results reveal a significant difference between Models 1 and 2 indicating slope bias for the M and A composites predicting FSG. Additionally, Models 2 and 3 were found to be significantly different for G and E predicting FSG. This indicates the presence of intercept bias for Air Traffic Control and Aircraft Electrical System trainees. For the TTC criterion, significant intercept differences were detected for the G and E composites. However, no slope or intercept bias was detected for M and TTC, although M was found to be a significant predictor of TTC. Finally, all of the  $F$  values for the linear models for A and TTC, including the test of model 3 against the null model, were nonsignificant.

Table 5 indicates the magnitude and direction of the level bias found for the G and E composites

---

Insert Table 5 about here

---

when predicting FSG and TTC. The scores in this table reflect the predicted criterion scores for subgroup members with equivalent aptitudes relative to predicted criterion scores derived from using a common regression line. The results show that using a common regression line for both the G and E composites will overpredict criterion scores for blacks for FSG. Similarly, a common regression line will "overpredict" blacks' criterion scores for TTC (given that lower TTC scores are more desirable than higher TTC scores). Moreover, the predicted scores derived from a common regression line were consistently closer to the predicted scores for whites than for blacks for both FSG and TTC. For example, for FSG, whites' scores were underpredicted by a common regression line by .27 grade points for the G composite, and .21 grade points for the E composite. For blacks, however, FSG scores were overpredicted by 2.05 grade points for the G composite, and .64 grade points for the E composite. A similar pattern emerged for the TTC criterion. The performance for whites was underpredicted by .61 days for the G composite, and 1.19 days for the E composite whereas black performance was overpredicted by 2.75 days by the G composite and 5.25 days for the E composite.

Figures 1 through 4 also illustrate the magnitude of the level bias effects. Each figure represents a plot of

---

Insert Figures 1-8 about here

---

the final linear model for each of the training courses. For the G and E composites and FSG, blacks' predicted scores were between 2.37 (.40 s.d.) to .83 grade points (.13 s.d.) below the predicted scores for whites, respectively. For the G and E composites predicting TTC, the differences between the scores for blacks and the scores for whites were slightly less pronounced. For the Air Traffic Control course, blacks were predicted to need approximately 3 more days (.12 s.d.) to complete training than whites. Blacks in the Aircraft Electrical System course were predicted to need 6.71 more training days (.24 s.d.) than whites.

Figures 5 and 7 indicate slope bias for the M and A composites predicting FSG. Figure 5 represents a disordinal interaction with the intercept for blacks 1.42 grade points *higher* (.24 s.d.) than the intercept for whites at the M cutoff score of 40. At the 100th percentile score for the M composite, the predicted FSG for blacks was 2.7 grade points (.46 s.d.) *lower* than the predicted FSG for whites. At the mean for the M composite ( $\bar{M} = 66.91$ ) the predicted score for whites was .5 grade points (.09 s.d.) higher than the predicted score for blacks.

Figure 7 shows the slope bias present for the A composite predicting FSG. This figure illustrates that at the cutoff score of 40 for the A composite, the intercept for whites was .74 grade points (.11 s.d.) higher than the intercept for blacks. At the 100th percentile the predicted score for whites was 3.2 grade points (.46 s.d.) higher than the predicted score for blacks. However, looking at the mean for the A composite ( $\bar{M} = 67.65$ ), the predicted score for whites was 2 grade points (.29 s.d.) higher than the predicted score for blacks. The use of a common regression line, however, helped to attenuate these differences. When a common regression line was used for this training course, the FSG for blacks was underpredicted by .25 (.04 s.d.) grade points at a score of 40 on the A composite, 1 grade point (.14 s.d.) at the mean of A, and 2.25 (.33 s.d.) grade points at the 100th percentile.

Figure 6 shows that the final regression line for the 2 subgroups is a common regression line. Figure 8 shows the lack of predictive validity of the A composite for TTC.

#### Discussion

The results of the present study are generally consistent with past research on the equity of the ASVAB and other measures of cognitive ability (Jensen, 1980; Sackett & Wilk, 1994; Welsh et al., 1990). First, hypothesis one was partially supported in that level bias was found for the G and E composites and FSG. Moreover, in both conditions the intercept for whites was higher than the intercept for blacks. Thus, if a common regression line were used to predict scores for both racial subgroups, FSG for blacks would be overpredicted. Contrary to hypothesis 1, however, slope bias was detected for the M and A composites predicting FSG.

Hypothesis 2 was also partially supported. Level bias was detected for the G and E composites predicting TTC. As with the level bias found with FSG, the use of a common regression line for TTC and the G and E composites would "overpredict" the scores for blacks. A common regression line would predict that blacks would require fewer days to complete training than would be predicted using a separate regression line for black airmen only. For the M composite and TTC, no level or slope bias was detected. However, contrary to hypothesis 2, the A composite was not found to be a valid predictor of TTC.

Hypothesis 3 received only limited support. For Air Traffic Control trainees, predicted FSG for blacks was .40 s.d. below the predicted FSG for whites, but blacks' predicted TTC was only .12 s.d. "below" whites' predicted scores. For Aircraft Electrical System trainees, however, this pattern was reversed. For this condition, there were smaller intercept differences for FSG (.13 s.d.) than for TTC (.24 s.d.). For Aerospace Ground Equipment trainees, no intercept or slope bias was detected for TTC, but slope bias was found for FSG. For Administrative Staff Support, the A composite was not found to be related to TTC. Although A was not predicted to be uncorrelated with TTC, the general pattern of results found for these two variables is consistent with a review of ASVAB validity studies done by Welsh et al. (1990). This review indicates that the lowest correlations are typically found with the A composite and the TTC criterion.

Although both level and slope bias were detected in the present study, the results of this study should be considered in the context of practical significance as well as statistical significance due to the large sample sizes employed. Practical significance is usually assessed by looking at the effect size for a statistically significant results (Hays, 1981). The effect sizes associated with the significance values for the training courses where slope and intercept bias were detected are presented in Table 4. As can be seen from this table, an  $R$ -square change value of .001 was found to be statistically significantly at  $p < .001$  for the A composite predicting FSG, and the G composite predicting TTC. Thus, the large sample sizes in each of the training courses analyzed in this study resulted in statistically significant differences between  $R$  squared change values that may not have been found to be statistically different if smaller sample sizes were used. When these results are considered in the larger context of practical significance, the results indicated that the predicted score differences between blacks and whites ranged from .12 to .46 of a standard deviation. However, when level bias was detected, the direction of the difference between the predicted scores was such that the use of a common regression line would overpredict FSG and TTC. In this instance, the test is biased according to the Cleary rule (1968) but the test is not unfair because more minorities would be selected by using a common regression line than separate regression lines for the subgroups (Sackett & Wilk, 1994). This pattern of results was found for two of the four training courses for both the FSG and TTC.

In addition to level bias, however, the present study also detected slope bias for FSG for two of the four training courses. Slope bias presents a more serious problem than level bias because the use of a common regression line may over or underpredict criterion scores for the minority group depending on the level of the predictor. This pattern of results was found for Aerospace Ground equipment trainees and FSG. Slope bias was also detected for Administrative Staff Support trainees. However, this interaction was not disordinal across the relevant range of the predictor scores. Thus, the use of a common regression line to predict FSG for this training course would always overpredict the scores for blacks, but the predicted FSG scores would vary across levels of the aptitude predictor from as little as .11 to as much as .46 s.d. below whites. Where slope bias is detected, separate regression lines should be used for the majority and minority subgroups (Cascio, 1989; Sackett & Wilk, 1994). Thus, while some of the actual differences between predicted scores for the subgroups appear trivial (e.g., E

composite predicting FSG revealed a .83 grade point difference in prediction between racial subgroups) others are more consequential.

In line with judging the results of the current study against practical versus just statistical significance, one interesting implication of this study is that the MGE composites can be used to predict the time it will take individuals to complete training. A report by the Automated Science Group and CACI (1988) estimated cost savings associated with validity increments for training criteria of .02 ( $R^2$ ) to be in excess of \$80 million per year across all four service branches. Although costs of training new recruits to replace individuals who failed training vary from Service to Service, (e.g., \$1,800 - \$4,300 per recruit in 1986) these costs capture only part of the cost associated with training and training losses. In short, the military spends billions of dollars on training new recruits. Consequently, the use of a valid and unbiased predictor to predict the number of training days required for individuals with a certain minimal aptitude level could translate "small" effect sizes into billions of dollars.

In sum, the present study expanded past research on the equity of the ASVAB in three ways. First, time to complete training was used as another training criterion in addition to the more often used criterion of FSG. Second, TTC was evaluated in the context of an equity study whereas past research has focused on TTC primarily in the context of validation studies (*cf.* Welsh et al., 1990). Finally, for two of the four training courses, the discrepancies between the scores for blacks and whites were smaller for TTC than for FSG. However, the lack of bias or smaller magnitude of bias for TTC may come at the price of lower average validities and higher standard errors for the TTC criterion. Although the predictor composites were found to be significantly correlated with TTC (with the exception of administrative aptitude), Welsh et al. (1990) suggest that the average validities for TTC are lower than FSG validities due to criterion contamination in the measure of TTC. They suggest that individuals with high ability may not be motivated to complete training as fast as possible. High ability trainees may not choose to complete a self-paced course quickly because they may have to "wait in the dormitories for the next block of instruction to begin" or they may be assigned to undesirable special details such as sweeping, cleaning, or grounds maintenance (Welsh et al., 1990). Additionally, TTC may be contaminated with rater bias much the same way FSG was hypothesized to be contaminated. Trainees must still reach some minimal level of proficiency to complete a self-paced training course. For the courses evaluated in the present study, instructors may have had the ability to determine how quickly or slowly a trainee progressed through training by requiring some trainees to repeat certain sections of the course. However, the degree to which TTC was contaminated with rater bias or other biases is not known due to this study's reliance on archival data spanning a period of over a decade. Future research should address the issue of comparing the potential bias for FSG and TTC in a more controlled environment to eliminate the potential for the contamination of TTC.

## References

- Bretz, R.D., Milkovich, G.T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. Journal of Management, 18(2), 321-352.
- Cascio, W.F. (1991). Applied psychology in personnel management (fourth edition). New Jersey: Prentice Hall.
- Cleary, T.A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 5, 114-124.
- Gordon, M.A. (1953). A study in the applicability of the same minimum qualifying scores for technical schools to white males, WAF and negro males (AFHRL-TR-53-34). Lackland, AFB, TX: Human Resources Research Center.
- Hays, W.L. (1988). Statistics (fourth edition). Forth Worth: Holt, Rinehart, and Winston, Inc.
- Hunter, J.E., Schmidt, F.L., & Rauchenberger, J.M. (1977). Differential validity of employment tests: Implications of four definitions for selection utility and minority hiring. Journal of Applied Psychology, 62, 245-260.
- Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.
- Kraiger, K., & Ford, J.K. (1985). A meta-analysis of ratee race effects in performance ratings. Journal of Applied Psychology, 70, 56-65.
- Maier, M.H., & Truss, A.R. (1984). Validity of the occupational and academic composites for the Armed Services Vocational Aptitude Battery, Form 14, in Marine Corps Training Courses. (Memorandum N. 84-3043/2). Alexandria, VA: Center for Naval Analyses.
- McLaughlin, D.H., Rossmeissl, P.G., Wise, L.L., Brandt, D.A., & Wang, M. (1984). Validation of current alternative Armed Services Vocational Aptitude Battery (ASVAB) area composites, based on training and Skill Qualification Test (SQT) information in fiscal year 1981 and 1982 (ARI-TR-651, AD-A156 807). Alexandria, VA: Army Research Institute.
- Nunnally, J.C., & Bernstein, I.H. (1994). Psychometric theory (third edition). New York: McGraw-Hill, Inc.
- Pedhazur, E.J. (1982). Multiple regression in behavioral research (second edition). New York: Holt, Rinehart and Winston, Inc.
- Rothstein, H.R., & McDaniel, M.A. (1991). Mental abilities testing in industry. In J.W. Jones, B.D. Steffuy, & D.W. Bray (Eds.) Applying Psychology in Business (pp. 236-242). Lexington, MA: Lexington Book.
- Sackett, P.R., & Wilk S.L. (1994). Within-group norming and other forms of score adjustment in pre-employment testing. American psychologist, 49(11), 929-954.
- Stone-Romero, E.F., Alliger, G.M., & Aguinis, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects of dichotomous variables. Journal of Management, 20(1), 167-178.



Vitola, B.M., & Alley, W.E. (1968). Development and standardization of Air Force composites for the Armed Services Vocational Aptitude Battery. (AFHRL-TR-68-110) Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Ward, J.H., & Jennings, E. (1973). Introduction to linear models. Englewood Cliffs, NJ: Prentice-Hall.

Welsh, J.R., Kucinkas, S.K., & Curran, L.T. (1990). Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies (AFHRL-TR-90-22). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Wise, L., Welsh, J., Grafton, F., Foley, P., Earles, J., Sawin, L., & Divgi, D.R. (1992). Sensitivity and fairness of the Armed Services Vocational Aptitude Battery (ASVAB) technical composites (DMDC Technical Report 92-002). Monterey, CA: Personnel Testing Division, Defense Manpower Data Center.

Wilbourn, J.M., Valentine, L.D., & Ree, M.J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9 and 10 to Air Force technical final school grades (AFHRL-TR-84-8, AD-A144 213). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.



Table 1

Means and Standard Deviations for Percentile Score Predictor and Criterion Variables

<u>Course</u>	<u>Total Group</u>					<u>Whites</u>					<u>Blacks</u>				
	<u>Apt.</u>	<u>N</u>	<u>Apt.</u>	<u>FSG</u>	<u>TTC</u>	<u>N</u>	<u>Apt.</u>	<u>FSG</u>	<u>TTC</u>	<u>N</u>	<u>Apt.</u>	<u>FSG</u>	<u>TTC</u>	<u>N</u>	<u>Apt.</u>
<u>Aerospace Ground Equipment</u>	M	3802	66.91 [17.56]	84.72 [5.88]	109.70 [14.73]	3476	68.11 [17.42]	84.82 [5.92]	109.40 [14.81]	326	54.08 [14.44]	83.66 [5.40]	112.4 [13.57]		
<u>Staff Support Administration</u>	A	11,728	67.65 [15.27]	85.58 [6.92]	43.17 [7.18]	7750	69.52 [15.30]	86.32 [6.91]	43.08 [7.20]	3978	63.98 [14.54]	84.14 [6.73]	43.36 [7.14]		
<u>Air Traffic Control</u>	G	6290	76.42 [11.31]	86.68 [5.9]	147.4 [24.59]	5556	77.08 [11.33]	87.02 [5.84]	146.9 [24.51]	734	71.38 [9.83]	84.09 [5.73]	151.4 [24.83]		
<u>Aircraft Electrical Systems</u>	E	4465	69.69 [12.57]	79.52 [6.56]	110.6 [27.44]	3675	71.06 [12.72]	79.96 [6.56]	109.1 [27.10]	790	63.29 [9.54]	77.49 [6.15]	117.4 [28.00]		

Note. Standard deviations are shown in parentheses.

Table 2

Uncorrected (and Corrected) Zero-Order Correlations of Separate Aptitude Predictor Variables  
with Performance Criteria

Course	Predictor	Whites			Blacks		
		N	FSG	TD	N	FSG	TD
Aerospace Ground Equipment	M	3476	.34**** [.45]	-.22**** [-.31]	326	.12* [.21]	-.13* [-.21]
Staff Support Administration	A	7750	.19**** [.31]	-0.001 [-.002]	3978	.10**** [.22]	-.03* [-.07]
Air Traffic Control	G	5556	.19**** [.34]	-.13**** [-.24]	734	.19**** [.38]	-.12 *** [-.25]
Aircraft Electrical Systems	E	3675	.41**** [.61]	-.09**** [-.15]	790	.34**** [.63]	-.08* [-.18]

\*  $p < .05$

\*\*  $p < .01$

\*\*\*  $p < .001$

\*\*\*\*  $p < .0001$

Table 3

Standard Error of Estimates (SEE) of AFOQT Composite Variables with Performance

Course	Final School Grade				Time to Complete Training			
	White SEE	Black SEE	SEE Difference	F value	White SEE	Black SEE	SEE Difference	F value
Aerospace Ground Equipment	5.61	5.37	0.24	1.09	14.47	13.48	0.99	1.15
Staff Support Administration	6.79	6.71	0.08	1.02	7.21	7.14	0.07	1.02
Air Traffic Control	5.73	5.62	0.11	1.04	24.31	24.66	0.35	1.03
Aircraft Electrical Systems	5.99	5.81	0.18	1.06	26.99	27.91	0.92	1.07

Note. All F values were found to nonsignificant,  $p > .05$ .

Table 4

## Regression Analyses Results for Black and White Trainees

Criterion	Course	Multiple Correlation Squared (R-Square)			Model Comparisons R Square Change	
		Model 1	Model 2	Model 3	Slope Bias F (1 vs. 2)	Level Bias F (2 vs. 3)
Final School Grade	Aerospace Ground Equipment (M)	0.110	0.110	0.110	8.50**	1.47
	Staff Support Administration (A)	0.048	0.047	0.032	20.76***	175.28***
	Air Traffic Control (G)	0.061	0.061	0.045	0.18	108.28***
	Aircraft Electrical Systems (E)	0.174	0.174	0.171	0.08	12.05***
Time to Complete Training	Aerospace Ground Equipment (M)	0.047	0.047	0.047	0.78	0.31
	Staff Support Administration (A)	0.000	0.000	0.000	2.53	3.13
	Air Traffic Control (G)	0.019	0.019	0.018	0.15	9.10***
	Aircraft Electrical Systems (E)	0.021	0.021	0.013	0.19	37.46***

\*  $p < .05$ \*\*  $p < .01$ \*\*\*  $p < .001$

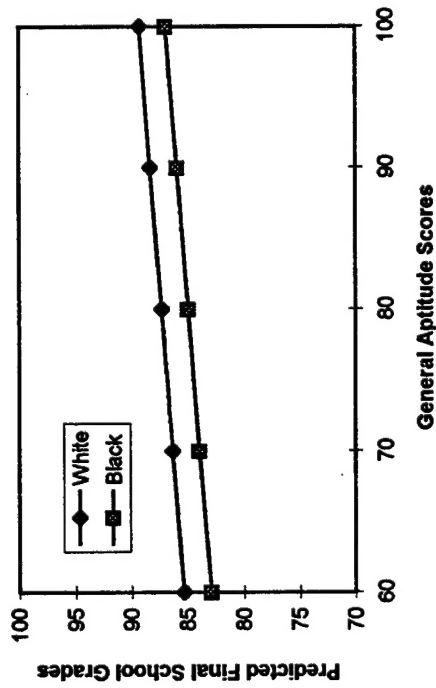
**Table 5**

**Magnitude and Direction of Level-Bias -- Predicted Performance Scores for Racial Subgroups**

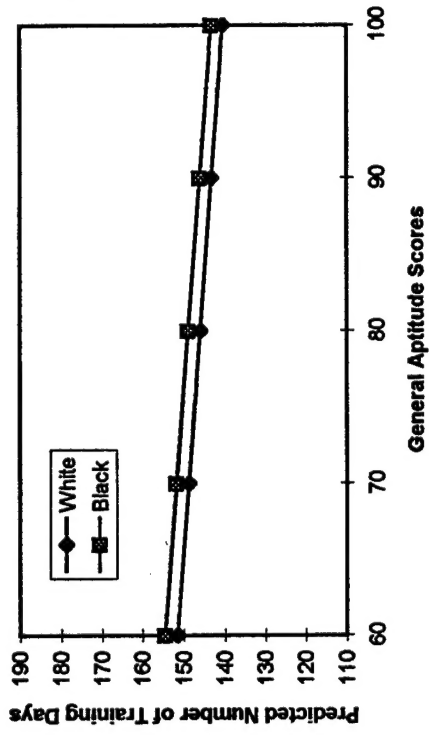
	Predicted Scores for White	Common Regression Line	Predicted Scores for Blacks
<hr/>			
<b>Final School Grade</b>			
General	86.95	86.68	84.63
Electronic	79.72	79.51	78.87
<hr/>			
<b>Time to Complete Training</b>			
General	151.61	152.22	154.97
Electronic	109.39	110.58	115.83
<hr/>			

**Note.** Mean Aptitude scores for the total subsamples were used to obtain mean predicted criterion scores for each subgroup and for a common regression line. The mean for the entire sample for General aptitude was 76.42, and the mean for the combined sample for Electronic aptitude was 69.69.

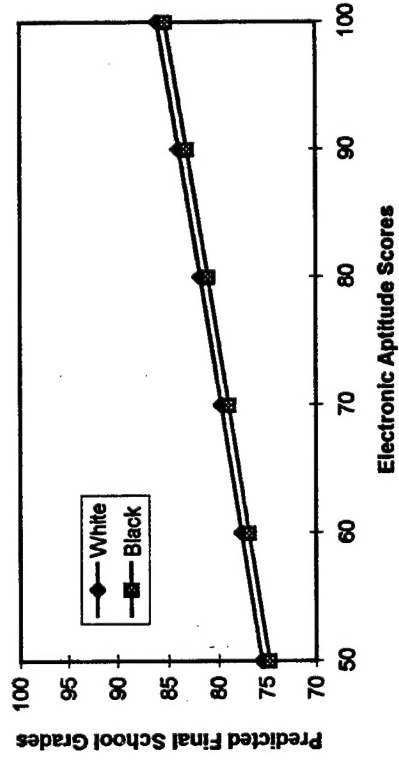
**Figure 1. Predicted Final School Grades for Air Traffic Control Trainees**



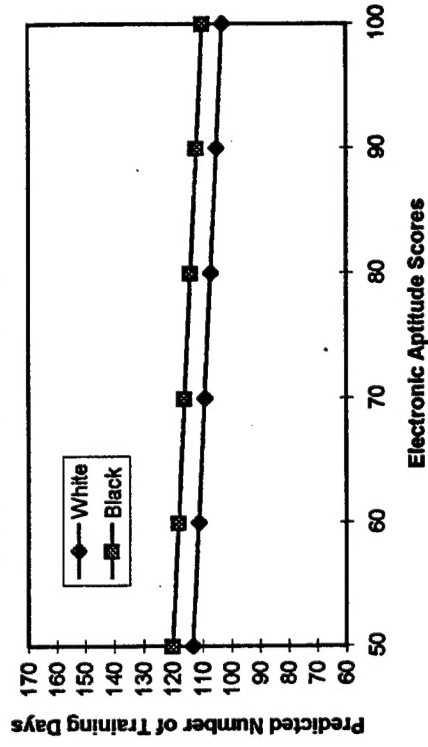
**Figure 2. Predicted Time in Training for Air Traffic Control Trainees**



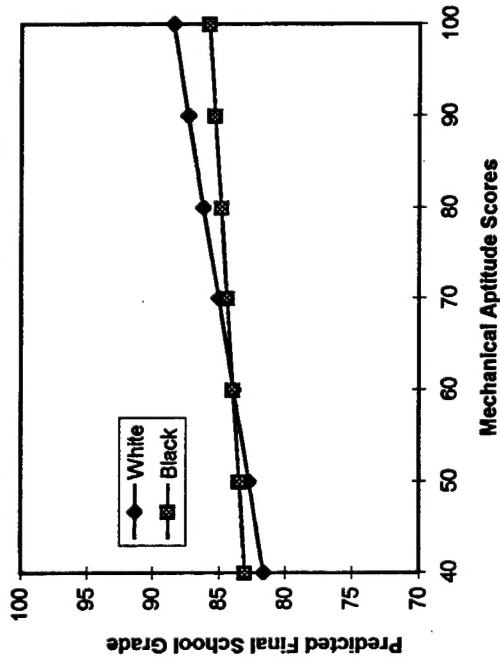
**Figure 3. Predicted Final School Grades for Aircraft Electrical System Trainees**



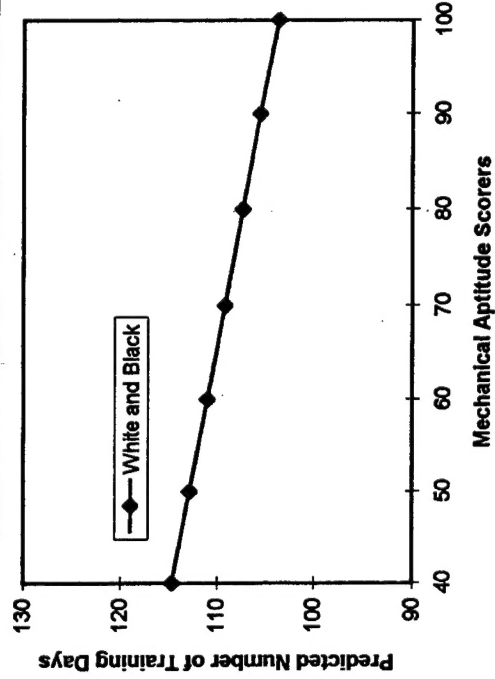
**Figure 4. Predicted Time in Training for Aircraft Electrical System Trainees**



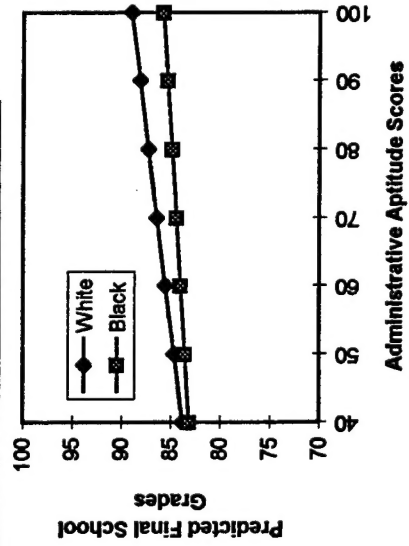
**Figure 5. Predicted Final School Grades for Aerospace Ground Equipment Trainees**



**Figure 6. Predicted Time to Complete Training for Aerospace Ground Equipment Trainees**



**Figure 7. Predicted Final School Grades for Administrative Staff Support Trainees**



**Figure 8. Predicted Time to Complete Training for Administrative Staff Support Trainees**

